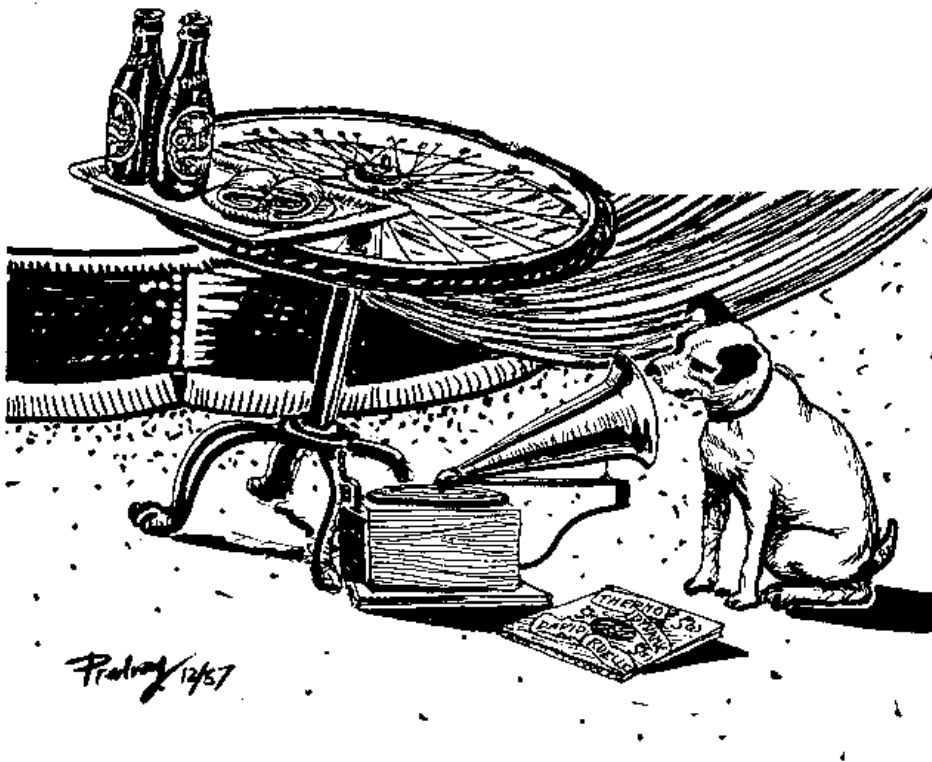


Chaos: Classical and Quantum



Predrag Cvitanović – Roberto Artuso – Ronnie Mainieri – Gregor Tanner –
Gábor Vattay

A very very short introduction

And if you don't know, now you know

— [The Notorious B.I.G.](#)

IF YOU UNDERSTAND the cycle averaging formulas (23.23) for the expectation of an observable in a chaotic flow, you don't need ChaosBook. If you don't understand, don't despair. No newborn understands the 2nd law, either.



ChaosBook is an advanced textbook on the theory of
deterministic, stochastic and quantum
chaotic / turbulent systems

on level of a 2nd year graduate physics statistical mechanics or quantum field theory course. Approach it the way that suits you best.

All this book says is that time evolution adds up probability densities of initial states. The problem is linear, so one solves it by finding its eigenvectors and eigenvalues, i.e., zeros of a determinant. This determinant is Greek to you, so they call it the 'zeta' function. A way to evaluate a determinant is in terms of traces. That they call the 'trace formula'. If we succeed, we will have taught you two things: (i) the difference (19.28) between a finite transformation and its generator, and (ii) that $\log \det = \text{tr} \log$. That's how you harness finite transformations (which might explode exponentially) in terms of integrals over generators (whose action is bounded), sect. 5.6 *Won't you be my neighbor?* and trace formula (21.19).

Now you know. So, have a look at the cycle averaging formulas (23.23), backtrack to where you have a surer footing, until you 'get' it.

Any novice can master ChaosBook part I *Geometry of chaos*, online course [part 1](#) - indeed, any scientist, engineer or mathematician would profit from understanding nonlinear world on this level.

The theory developed in ChaosBook part II *Chaos rules* is here to challenge a seasoned theorist. She might start with chapter 21 *Trace formulas*, online course [part 2](#), and work her way back or forth, as needed.

Predrag Cvitanović, Atlanta, September 2024

Contents

A very very short introduction	i
Contents	xvii
Acknowledgments	xvii
 I Geometry of chaos	 1
1 Overture	2
1.1 Why ChaosBook?	3
1.2 Chaos ahead	4
1.3 The future as in a mirror	5
1.4 A game of pinball	11
1.5 Chaos for cyclists	16
1.6 Change in time	21
1.7 To statistical mechanics	24
1.8 Chaos: what is it good for?	25
1.9 What is not in ChaosBook	28
résumé 28 commentary 30 guide to exercises 34	
References	35
1.10 *	38
exercises 38	
 2 Go with the flow	 39
2.1 Dynamical systems	39
2.2 Flows	44
2.3 Changing coordinates	48
2.4 Life in extreme dimensions	50
2.5 Computing trajectories	55
résumé 55 commentary 56	
References	59
2.6 Examples	62
2.7 *	64
exercises 64	
 3 Discrete time dynamics	 68
3.1 Poincaré sections	69
3.2 Computing a Poincaré section	74
3.3 Mappings	75

résumé 77	commentary 78	
References		79
3.4 Examples		81
3.5 *		84
exercises	84	
4 Local stability		86
4.1 Flows transport neighborhoods		86
4.2 Computing the Jacobian matrix		89
4.3 A linear diversion		90
4.4 Stability of flows		92
4.5 Stability of maps		93
4.6 Stability of return maps		93
4.7 Neighborhood volume		95
résumé 96	commentary 97	
References		99
4.8 Examples		101
4.9 Exercises		107
exercises	107	
5 Cycle stability		109
5.1 Equilibria		110
5.2 Periodic orbits		110
5.3 Floquet multipliers are invariant		113
5.4 Floquet multipliers are metric invariants		115
5.5 Stability of return map cycles		117
5.6 Won't you be my neighbor?		117
résumé 118	commentary 119	
References		120
5.7 Examples		122
5.8 Exercises		123
exercises	123	
6 Lyapunov exponents		124
6.1 Stretch, strain and twirl		125
6.2 Lyapunov exponents		126
résumé 129	commentary 129	
References		132
6.3 Examples		133
6.4 Exercises		135
exercises	135	
7 Fixed points		136
7.1 One-dimensional maps		137
7.2 Flows		139
résumé 141	commentary 141	
References		141
7.3 Examples		142
7.4 *		144

exercises 144	
8 Hamiltonian dynamics	146
8.1 Hamiltonian flows	147
8.2 Hamiltonian maps	148
8.3 Symplectic group	150
8.4 Stability of Hamiltonian flows	151
8.5 Symplectic maps	153
8.6 Poincaré invariants	154
résumé 155 commentary 155	
References	158
8.7 Examples	161
8.8 *	166
exercises 166	
9 Billiards	168
9.1 Billiard dynamics	168
9.2 Stability of billiards	170
résumé 172 commentary 172	
References	173
9.3 Examples	175
9.4 *	176
exercises 176	
10 Flips, slides and turns	178
10.1 Discrete symmetries	179
10.2 Subgroups, cosets, classes	182
10.3 Orbits, quotient space	183
résumé 186 commentary 186	
References	186
10.4 Examples	188
10.5 Exercises	191
exercises 191	
11 World in a mirror	192
11.1 Symmetries of solutions	193
11.2 Relative periodic orbits	195
11.3 Dynamics reduced to fundamental domain	196
11.4 Life on the border	198
11.5 Invariant polynomials	199
résumé 200 commentary 200	
References	202
11.6 Examples	204
11.7 Exercises	209
exercises 209	

12 Relativity for cyclists	212
12.1 Continuous symmetries	213
12.2 Symmetries of solutions	218
12.3 Stability	223
résumé 224 commentary 225	
References	225
12.4 Examples	228
12.5 Exercises	233
exercises 233	
13 Slice & dice	235
13.1 Moving frames	236
13.2 Symmetry reduction	237
13.3 Bringing it all back home: method of slices	237
13.4 Dynamics within a slice	240
13.5 First Fourier mode slice	242
13.6 Method of images: Hilbert bases	244
résumé 246 commentary 247	
References	251
13.7 Examples	256
13.8 *	258
exercises 258	
14 Charting the state space	260
14.1 Qualitative dynamics	261
14.2 Stretch and fold	264
14.3 Temporal ordering: Itineraries	266
14.4 Spatial ordering	268
14.5 Kneading theory	270
14.6 Symbolic dynamics, basic notions	272
résumé 275 commentary 276	
References	278
14.7 Examples	282
14.8 *	288
exercises 288	
15 Stretch, fold, prune	290
15.1 Goin' global: stable/unstable manifolds	291
15.2 Horseshoes	295
15.3 Symbol plane	297
15.4 Prune Danish	300
15.5 Recoding, symmetries, tilings	303
15.6 Charting the state space	305
résumé 308 commentary 309	
References	310
15.7 Examples	314
15.8 *	318
exercises 318	

16 Fixed points, and how to get them	321
16.1 Where are the cycles?	322
16.2 Multipoint shooting method	324
16.3 Cost function	326
résumé 328 commentary 328	
References	329
16.4 Examples	330
16.5 *	332
exercises 332	
 II Chaos rules	 334
17 Walkabout: Transition graphs	335
17.1 Matrix representations of topological dynamics	335
17.2 Transition graphs: wander from node to node	337
17.3 Transition graphs: stroll from link to link	338
résumé 341 commentary 341	
References	342
17.4 Examples	344
17.5 *	348
exercises 348	
 18 Counting	 349
18.1 How many ways to get there from here?	350
18.2 Topological trace formula	351
18.3 Determinant of a graph	354
18.4 Topological zeta function	356
18.5 Infinite partitions	358
18.6 Shadowing	360
18.7 Counting cycles	361
résumé 363 commentary 365	
References	366
18.8 Examples	368
18.9 *	376
exercises 376	
 19 Transporting densities	 381
19.1 Measures	382
19.2 Perron-Frobenius operator	383
19.3 Why not just leave it to a computer?	385
19.4 Invariant measures	387
19.5 Density evolution for infinitesimal times	391
19.6 Liouville operator	392
résumé 393 commentary 394	
References	395
19.7 Examples	398
19.8 Exercises	399
exercises 399	

20 Averaging	402
20.1 Dynamical averaging	402
20.2 Evolution operators	406
20.3 Averaging in open systems	411
20.4 Evolution operator evaluation of Lyapunov exponents	412
résumé 413 commentary 414	
References	414
20.5 Examples	415
20.6 *	418
exercises 418	
21 Trace formulas	419
21.1 A trace formula for maps	420
21.2 A trace formula for flows	423
21.3 An asymptotic trace formula	427
résumé 428 commentary 428	
References	429
21.4 Examples	430
21.5 *	432
exercises 432	
22 Spectral determinants	433
22.1 Spectral determinants for maps	433
22.2 Spectral determinant for flows	434
22.3 Dynamical zeta functions	436
22.4 False zeros	437
22.5 Spectral determinants vs. dynamical zeta functions	438
22.6 All too many eigenvalues?	438
résumé 439 commentary 440	
References	442
22.7 Examples	444
22.8 *	446
exercises 446	
23 Cycle expansions	448
23.1 Pseudo-cycles and shadowing	449
23.2 Construction of cycle expansions	451
23.3 Periodic orbit averaging	456
23.4 Flow conservation sum rules	457
23.5 Cycle formulas for dynamical averages	458
23.6 Cycle expansions for finite alphabets	461
23.7 Stability ordering of cycle expansions	462
résumé 464 commentary 465	
References	469
23.8 Examples	471
23.9 *	472
exercises 472	

24 Deterministic diffusion	475
24.1 Diffusion in periodic arrays	476
24.2 Diffusion induced by chains of 1-dimensional maps	480
24.3 Marginal stability and anomalous diffusion	483
résumé 486 commentary 487	
References	489
24.4 Examples	492
24.5 Exercises	497
exercises 497	
25 Discrete symmetry factorization	499
25.1 Transformation of functions	500
25.2 Taking care of fundamentals	501
25.3 Dynamics in the fundamental domain	504
25.4 Discrete symmetry factorizations	506
25.5 $Z_2 = D_1$ factorization	510
25.6 D_3 factorization: 3-disk game of pinball	511
résumé 513 commentary 514	
References	515
25.7 Examples	517
25.8 Exercises	523
exercises 523	
26 Continuous symmetry factorization	525
26.1 Compact groups	526
26.2 Continuous symmetries of dynamics	530
26.3 Symmetry reduced trace formula for flows	532
résumé 533 commentary 534	
References	536
26.4 Examples	538
26.5 Exercises	541
exercises 541	
III Chaos: what to do about it?	542
27 Why cycle?	543
27.1 Escape rates	543
27.2 Natural measure in terms of periodic orbits	545
27.3 Correlation functions	547
27.4 Trace formulas vs. level sums	548
résumé 550 commentary 550	
References	551
27.5 Exercises	553
exercises 553	

28 Why does it work?	555
28.1 Linear maps: exact spectra	556
28.2 Evolution operator in a matrix representation	558
28.3 Deterministic Fredholm theory	560
28.4 Analyticity of spectral determinants	562
28.5 Hyperbolic maps	566
28.6 Physics of eigenvalues and eigenfunctions	567
28.7 Troubles ahead	569
résumé 570 commentary 572	
References	573
28.8 Examples	575
28.9 Exercises	579
exercises 579	
29 Intermittency	580
29.1 Intermittency everywhere	581
29.2 Intermittency for pedestrians	583
29.3 Intermittency for cyclists	594
29.4 BER zeta functions	600
résumé 602 commentary 602	
References	603
29.5 Examples	606
29.6 Exercises	607
exercises 607	
30 Turbulence?	609
30.1 Configuration space: a fluttering flame front	609
30.2 Constructing a state space	612
30.3 Energy budget	617
30.4 Infinite-dimensional flows: Numerics	619
30.5 Visualization	621
30.6 Why does a flame front flutter?	621
30.7 Intrinsic parametrization	623
résumé 625 commentary 626	
References	627
30.8 Examples	631
30.9 Exercises	635
exercises 635	
31 Koopman modes	638
31.1 Koopmania	638
31.2 Koopman eigenvalues for a limit cycle	640
commentary 644	
References	644
31.3 Examples	645
31.4 *	648
exercises 648	

32 Irrationally winding	649
32.1 Mode locking	650
32.2 Local theory: "Golden mean" renormalization	655
32.3 Global theory: Thermodynamic averaging	656
32.4 Hausdorff dimension of irrational windings	658
32.5 Thermodynamics of Farey tree: Farey model	659
résumé 662 commentary 662	
References	665
32.6 Exercises	669
exercises 669	
 IV The rest is noise	 670
33 Noise	671
33.1 Deterministic transport	672
33.2 Brownian diffusion	673
33.3 Noisy trajectories	675
33.4 Noisy maps	678
33.5 All nonlinear noise is local	680
33.6 Weak noise: Hamiltonian formulation	682
résumé 684 commentary 684	
References	686
33.7 Examples	689
33.8 *	690
exercises 690	
 34 Relaxation for cyclists	 695
34.1 Fictitious time relaxation	696
34.2 Discrete iteration relaxation method	699
34.3 Least action method	701
résumé 703 commentary 703	
References	705
34.4 Examples	707
34.5 *	710
exercises 710	
 V Quantum chaos	 711
35 Prologue	712
35.1 Quantum pinball	713
35.2 Quantization of helium	715
commentary 716	
References	716

36 Quantum mechanics	718
- the short short version	
36.1 Exercises	721
exercises 721	
37 WKB quantization	725
37.1 WKB ansatz	725
37.2 Method of stationary phase	728
37.3 WKB quantization	728
37.4 Beyond the quadratic saddle point	731
résumé 732 commentary 732	
References	733
37.5 *	734
exercises 734	
38 Semiclassical evolution	735
38.1 Hamilton-Jacobi theory	735
38.2 Semiclassical propagator	743
38.3 Semiclassical Green's function	746
résumé 751 commentary 752	
References	753
38.4 Exercises	755
exercises 755	
39 Semiclassical quantization	757
39.1 Trace formula	757
39.2 Semiclassical spectral determinant	762
39.3 One-degree of freedom systems	763
39.4 Two-degrees of freedom systems	764
résumé 765 commentary 766	
References	767
39.5 Exercises	769
exercises 769	
40 Quantum scattering	770
40.1 Density of states	770
40.2 Quantum mechanical scattering matrix	774
40.3 Krein-Friedel-Lloyd formula	775
40.4 Wigner time delay	778
commentary 779	
References	780
40.5 *	782
exercises 782	
41 Chaotic multiscattering	783
41.1 Quantum mechanical scattering matrix	784
41.2 N -scatterer spectral determinant	787
41.3 Semiclassical 1-disk scattering	791
41.4 From quantum cycle to semiclassical cycle	797

41.5 Heisenberg uncertainty	799
commentary 800	
References	801
42 Helium atom	803
42.1 Deterministic dynamics of collinear helium	804
42.2 Chaos, symbolic dynamics and periodic orbits	805
42.3 Local coordinates, Jacobian matrix	810
42.4 Getting ready	811
42.5 Semiclassical quantization of collinear helium	812
résumé 818 commentary 819	
References	821
42.6 Exercises	823
exercises 823	
43 Diffraction distraction	825
43.1 Quantum eavesdropping	825
43.2 An application	831
résumé 835 commentary 836	
References	837
43.3 Exercises	839
exercises 839	
Epilogue	840
Contributors	845
Index	849
VI Web appendices	871

A1 A brief history of chaos	872
A1.1 Chaos is born	872
A1.2 Chaos grows up	876
A1.3 Chaos with us	877
A1.4 Periodic orbit theory	879
A1.5 Dynamicist's vision of turbulence	885
A1.6 Gruppenpest	889
A1.7 Death of the Old Quantum Theory	891
commentary	894
References	898
A2 Go straight	909
A2.1 Rectification of flows	909
A2.2 Collinear helium	910
A2.3 Rectification of maps	914
A2.4 Rectification of a periodic orbit	915
résumé	916
commentary	917
References	918
A2.5 Examples	919
A2.6 Exercises	921
exercises	921
A4 Linear stability	922
A4.1 Linear algebra	923
A4.2 Eigenvalues and eigenvectors	925
A4.3 Eigenspectra: what to make out of them?	930
commentary	931
References	932
A4.4 Examples	933
A4.5 *	936
exercises	936
A6 Lyapunov exponents	937
A6.1 Evolution operator for Lyapunov exponents	937
A6.2 Advection of vector fields by chaotic flows	941
commentary	944
References	945
A6.3 *	946
exercises	946
A8 Hamiltonian dynamics	947
A8.1 Stability of Hamiltonian flows	947
A8.2 Monodromy matrix for Hamiltonian flows	948
A10 Flips, slides and turns	951
A10.1 Preliminaries and definitions	951
A10.2 Invariants and reducibility	957
A10.3 Hénon map symmetries	961
References	961

A10.4	Examples	962
A10.5	Exercises	965
	exercises	965
A14	Charting the state space	968
A14.1	Periodic orbits of unimodal maps	968
A14.2	Unimodal map bifurcation sequences	969
A14.3	Pruned Bernoulli shift	970
	commentary	975
	References	976
A16	Finding cycles	977
A16.1	Newton-Raphson method	977
A16.2	Hybrid Newton-Raphson / relaxation method	978
A18	Counting	981
A18.1	Topological zeta functions for infinite subshifts	981
A18.2	Prime factorization for dynamical itineraries	987
A18.3	Counting curvatures	990
	commentary	991
	References	992
A18.4	Exercises	993
	exercises	993
A20	Averaging	994
A20.1	Moments, cumulants	994
	commentary	997
	References	998
A20.2	Examples	999
A20.3	Exercises	1005
	exercises	1005
A22	Spectral determinants	1007
A22.1	Transfer operators	1007
	commentary	1011
	References	1012
A22.2	Examples	1012
A22.3	Exercises	1013
	exercises	1013
A24	Deterministic diffusion	1014
A24.1	Wanderings of a drunken snail	1014
A24.2	Lattice derivatives	1021
A24.3	Periodic lattices	1025
A24.4	Discrete Fourier transforms	1025
A24.5	Continuum field theory	1031
A24.6	Diffusion in sawtooth and cat maps	1032
	commentary	1036
	References	1037

A24.7Examples	1039
A24.8*	1042
exercises 1042	
A25Discrete symmetry factorization	1048
A25.1D ₄ factorization	1048
A25.2D ₂ factorization	1052
commentary 1054	
References	1055
A25.3Examples	1056
A31Koopman modes	1058
A31.1Koopmania	1058
A31.2Implementing evolution	1059
commentary 1061	
References	1061
A31.3*	1062
exercises 1062	
A32Thermodynamic formalism	1063
A32.1Rényi entropies	1063
A32.2Fractal dimensions	1067
résumé 1070 commentary 1071	
References	1071
A32.3Exercises	1073
exercises 1073	
A33Statistical mechanics recycled	1075
A33.1The thermodynamic limit	1075
A33.2Ising models	1077
A33.3Fisher droplet model	1080
A33.4Scaling functions	1085
A33.5Geometrization	1088
résumé 1095 commentary 1095	
References	1096
A33.6Exercises	1097
exercises 1097	
A37WKB quantization	1099
A37.1Semiclassical field theory, a single lattice site	1100
A37.2Semiclassical lattice field theory	1100
résumé 1101 commentary 1102	
References	1102
A39Semiclassical quantization, with corrections	1104
A39.1Periodic orbits as integrable systems	1104
A39.2The Birkhoff normal form	1107
A39.3Bohr-Sommerfeld quantization of periodic orbits	1109
A39.4Quantum calculation of $\hbar$ corrections	1110

commentary	1116
References	1116
A39.5 Examples	1117
A40 Infinite dimensional operators	1119
A40.1 Matrix-valued functions	1120
A40.2 Operator norms	1121
A40.3 Trace class and Hilbert-Schmidt class	1122
A40.4 Determinants of trace class operators	1124
A40.5 Von Koch matrices	1128
A40.6 Regularization	1129
References	1131
A40.7*	1132
exercises	1132
A41 Projects	1133

Acknowledgments

I feel I never want to write another book. What's the good!
I can eke living on stories and little articles, that don't cost
a tithe of the output a book costs. Why write novels any
more!

—D.H. Lawrence

This book owes its existence to the Niels Bohr Institute's and Nordita's hospitable and nurturing environment, and the private, national and cross-national foundations that have supported the collaborators' research over a span of several decades. P.C. thanks M.J. Feigenbaum of Rockefeller University; D. Ruelle of I.H.E.S., Bures-sur-Yvette; I. Procaccia of Minerva Center for Nonlinear Physics of Complex Systems, Weizmann Institute of Science; P.H. Damgaard of the Niels Bohr International Academy; G. Mazenko of U. of Chicago James Franck Institute and Argonne National Laboratory; T. Geisel of Max-Planck-Institut für Dynamik und Selbstorganisation, Göttingen; I. Andrić of Rudjer Bošković Institute; P. Hemmer of University of Trondheim; The Max-Planck Institut für Mathematik, Bonn; J. Lowenstein of New York University; Edificio Celi, Milano; Fundação de Faca, Porto Seguro; and Dr. Dj. Cvitanović, Kostrena, for the hospitality during various stages of this work, and the Carlsberg Foundation, Glen P. Robinson, Humboldt Foundation and National Science Foundation grant DMS-0807574 for partial support.

The authors gratefully acknowledge collaborations and/or stimulating discussions with E. Aurell, M. Avila, V. Baladi, D. Barkley, B. Brenner, G. Byrne, A. de Carvalho, D.J. Driebe, B. Eckhardt, M.J. Feigenbaum, J. Frøjlund, S. Froehlich, P. Gaspar, P. Gaspard, J. Guckenheimer, G.H. Gunaratne, P. Grassberger, H. Gutowitz, M. Gutzwiller, K.T. Hansen, P.J. Holmes, T. Janssen, R. Klages, T. Kreilos, Y. Lan, B. Lauritzen, C. Marcotte, J. Milnor, M. Nordahl, I. Procaccia, J.M. Robbins, P.E. Rosenqvist, D. Ruelle, G. Russberg, B. Sandstede, A. Shapere, M. Sieber, D. Sullivan, N. Søndergaard, T. Tél, C. Tresser, R. Wilczak, and D. Wintgen.

We thank Dorte Glass, Tzatzilha Torres Guadarrama and Raenell Soller for typing parts of the manuscript; D. Borrero, P. Düren, B. Lautrup, J.F. Gibson, M. Gomilšek and D. Viswanath for comments and corrections to the preliminary versions of this text; M.A. Porter for patiently and critically reading the manuscript, and then lengthening by the 2013 definite articles hitherto missing; M.V. Berry for the quotation on page 872; H. Fogedby for the quotation on page 562; J. Greensite for the quotation on page 6; S. Ortega Arango for the quotation on page 16; Ya.B. Pesin for the remarks quoted on page 896; M.A. Porter for the quotations on pages 8.1, 20, 16, 1.6 and A1.4; E.A. Spiegel for quotation on page 2; and E. Valesco for the quotation on page 25.

F. Haake's heartfelt lament on page 423 was uttered at the end of the first conference presentation of cycle expansions, November 1988. G.P. Morriss advice to students as how to read the introduction to this book, page 5, was offered during a 2002 graduate course in Dresden. J. Bellissard's advice to students concerning unpleasant operators and things nonlinear, pages 4.3 and 20.2.1, was shared in his 2013 "Classical Mechanics II" lectures on manifolds. K. Huang's C.N. Yang interview quoted on page 388 is available on ChaosBook.org/extras. T.D. Lee remarks on as to who is to blame, page 39 and page 322, as well as M. Shub's

helpful technical remark on page 572 came during the Rockefeller University December 2004 “Feigenbaum Fest.” Quotes on pages 39, 146, and 385 are taken from a book review by J. Guckenheimer [1].

Who is the 3-legged dog reappearing throughout the book? Long ago, when we were innocent and knew not Borel measurable α to Ω sets, P. Cvitanović asked V. Baladi a question about dynamical zeta functions, who then asked J.-P. Eckmann, who then asked D. Ruelle. The answer was transmitted back: “The master says: ‘It is holomorphic in a strip’.” Hence His Master’s Voice logo, and the 3-legged dog is us, still eager to fetch the bone. The answer has made it to the book, though not precisely in His Master’s voice. As a matter of fact, the answer *is* the book. We are still chewing on it.

What about the two beers? During his PhD studies, R. Artuso found the smørrebrød at the Niels Bohr Institute indigestible, so he digested H.M.V.’s wisdom on a strict diet of two Carlsbergs and two pieces of danish pastry for lunch every day, as depicted on the cover. Frequent trips back to Milano family kept him alive—he never got desperate enough to try the Danish smørrebrød. And the cycle wheel? Well, this is no book for pedestrians.

And last but not least: profound thanks to all the unsung heroes –students and colleagues, too numerous to list here– who have supported this project over many years in many ways, by surviving pilot courses based on this book, by providing invaluable insights, by teaching us, by inspiring us.

Part I

Geometry of chaos

WE START OUT with a recapitulation of the basic notions of dynamics. Our aim is narrow; we keep the exposition focused on prerequisites to the applications to be developed in this text. We assume that the reader is familiar with dynamics on the level of the introductory texts mentioned in remark 1.1, and concentrate here on developing intuition about what a dynamical system can do. It will be a broad stroke description, since describing all possible behaviors of dynamical systems is beyond human ken. While for a novice there is no shortcut through this lengthy detour, a sophisticated traveler might bravely skip this well-trodden territory and embark upon the journey at chapter 18.

The fate has handed you a law of nature. What are you to do with it?

1. Define your *dynamical system* (M, f) : the space M of its possible states, and the law f^t of their evolution in time.
2. Pin it down locally—is there anything about it that is stationary? Try to determine its *equilibria / fixed points* (chapter 2).
3. Cut across it, represent as a return map from a section to a section (chapter 3).
4. Explore the neighborhood by *linearizing* the flow; check the *linear stability* of its equilibria / fixed points, their stability eigen-directions (chapters 4 and 5).
5. Does your system have a *symmetry*? If so, you must use it (chapters 10 to 12). Slice & dice it (chapter 13).
6. Go global: train by *partitioning the state space* of 1-dimensional maps. Label the regions by *symbolic dynamics* (chapter 14).
7. Now venture global distances across the system by continuing local tangent space into *stable / unstable manifolds*. Their intersections *partition the state space* in a dynamically invariant way (chapter 15).
8. Guided by this topological partition, compute a set of *periodic orbits* up to a given topological length (chapter 7 and chapter 16).

Along the way you might want to learn about Lyapunov exponents (chapter 6), deterministic mechanics (chapter 8), and billiards (chapter 9).

Chapter 1

Overture

If I have seen less far than other men it is because I have stood behind giants.

—Eduardo Specchio

REREADING classic theoretical physics textbooks leaves a sense that there are holes large enough to steam a Eurostar train through them. Here we learn about harmonic oscillators and Keplerian ellipses - but where is the chapter on chaotic oscillators, the tumbling Hyperion? We have just quantized hydrogen, where is the chapter on the deterministic 3-body problem and its implications for quantization of helium? We have learned that an instanton is a solution of field-theoretic equations of motion, but shouldn't a strongly nonlinear field theory have turbulent solutions? How are we to think about systems where things fall apart; the center cannot hold; every trajectory is unstable?

This chapter offers a quick survey of the main topics covered in the book. Throughout the book



indicates that the section is on a pedestrian level - you are expected to know/learn this material



indicates that the section is on a somewhat advanced, cyclist level



indicates that the section requires a hearty stomach and is probably best skipped on first reading



fast track points you where to skip to



tells you where to go for more depth on a particular topic



link to a related video



[exercise 1.2] on margin links to an exercise that might clarify a point in the text



indicates that a figure is still missing—you are urged to fetch it

We start out by making promises—we will right wrongs, no longer shall you suffer the slings and arrows of outrageous Science of Perplexity. We relegate a historical overview of the development of chaotic dynamics to appendix [A1](#), and head straight to the starting line: A pinball game is used to motivate and illustrate most of the concepts to be developed in ChaosBook.

This is a textbook, not a research monograph, and you should be able to follow the thread of the argument without constant excursions to sources. Hence there are no literature references in the text proper, all learned remarks and bibliographical pointers are relegated to the “Commentary” section at the end of each chapter.

[ChaosBook.org](#) is continuously updated: recommended is always the latest ‘unstable’ [ChaosBook.org/chapters/ChaosBook.pdf](#) hyperlinked edition. If you cite ChaosBook, as any honorable reader would, cite the current ‘stable’ edition.

1.1 Why ChaosBook?

It seems sometimes that through a preoccupation with science, we acquire a firmer hold over the vicissitudes of life and meet them with greater calm, but in reality we have done no more than to find a way to escape from our sorrows.

—Hermann Minkowski in a letter to David Hilbert

The problem has been with us since Newton’s first frustrating (and unsuccessful) crack at the 3-body problem, lunar dynamics. Nature is rich in systems governed by simple deterministic laws whose asymptotic dynamics are complex beyond belief, systems which are locally unstable (almost) everywhere but globally recurrent. How do we describe their long term dynamics?

The answer turns out to be that we have to evaluate a determinant, take a logarithm. It would hardly merit a learned treatise, were it not for the fact that this determinant that we are to compute is fashioned out of infinitely many infinitely small pieces. The feel is of statistical mechanics, and that is how the problem was solved; in the 1960’s the pieces were counted, and in the 1970’s they were weighted and assembled in a fashion that in beauty and in depth ranks along with thermodynamics, partition functions and path integrals amongst the crown jewels of theoretical physics.

This book is *not* a book about periodic orbits. The red thread throughout the text is the duality between the local, topological, short-time dynamically invariant compact sets (equilibria, periodic orbits, partially hyperbolic invariant tori) and the global long-time evolution of densities of trajectories. Chaotic dynamics is generated by the interplay of locally unstable motions, and the interweaving of their global stable and unstable manifolds. These features are robust and accessible in systems as noisy as slices of rat brains. Poincaré, the first to understand deterministic chaos, already said as much (modulo rat brains). Once this topology

is understood, a powerful theory yields the observable consequences of chaotic dynamics, such as atomic spectra, transport coefficients, turbulent shapes.

That is what we will focus on in ChaosBook. The book is a self-contained graduate textbook on deterministic and quantum chaos. Your professor does not know this material, so you are on your own. We will teach you how to evaluate a determinant, take a logarithm—stuff like that. Ideally, this should take 100 pages or so. Well, we fail—so far we have not found a way to traverse this material in less than a semester, or 200-300 page subset of this text. [Nothing to be done.](#)

Question 1.1. Professor K. Zweistein asks

Q Perhaps it is painfully obvious to the experts, but I have so far failed to find what I need by a hyperlink-assisted walk through ChaosBook. Shouldn't the textbook be clear about this? At present, the barrier to entry (having to read ChaosBook entirely, cover to cover) appears too steep for the working scientists to learn. Perhaps a simple illustrative example? paper? would help...

A OK, Karen.

1.2 Chaos ahead

Things fall apart; the centre cannot hold.

—W.B. Yeats, *The Second Coming*

The study of chaotic dynamics is no recent fashion. It did not start with the widespread use of the personal computer. Chaotic systems have been studied for over 200 years. During this time many have contributed, and the field followed no single line of development; rather one sees many interwoven strands of progress.

In retrospect many triumphs of both deterministic and quantum physics were a stroke of luck: a few integrable problems, such as the harmonic oscillator and the Kepler problem, though 'non-generic', have gotten us very far. The success has lulled us into a habit of expecting simple solutions to simple equations—an expectation tempered by our recently acquired ability to numerically scan the state space of non-integrable dynamical systems. The initial impression might be that all of our analytic tools have failed us, and that the chaotic systems are amenable only to numerical and statistical investigations. Nevertheless, a beautiful theory of deterministic chaos, of predictive quality comparable to that of the traditional perturbation expansions for nearly integrable systems, already exists.

In the traditional approach the integrable motions are used as zeroth-order approximations to physical systems, and weak nonlinearities are then accounted for perturbatively. For strongly nonlinear, non-integrable systems such expansions fail completely; at asymptotic times the dynamics exhibit amazingly rich structure which is not at all apparent in the integrable approximations. However, hidden in this apparent chaos is a rigid skeleton, a self-similar tree of *cycles* (periodic orbits) of increasing periods. The insight of the modern dynamical systems theory

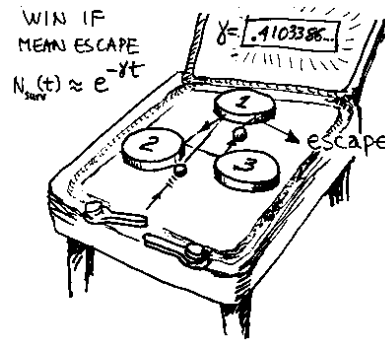


Figure 1.1: A physicist's bare bones game of pinball.

is that the zeroth-order approximations to the harshly chaotic dynamics should be very different from those for the nearly integrable systems: a good starting approximation here is the stretching and folding of baker's dough, rather than the periodic motion of a harmonic oscillator.

So, what is chaos, and what is to be done about it? To get some feeling for how and why unstable cycles come about, we start by playing a game of pinball. The remainder of the chapter is a quick tour through the material covered in *ChaosBook*. Do not worry if you do not understand every detail at the first reading—the intention is to give you a feeling for the main themes of the book. Details will be filled out later. If you want to get a particular point clarified right now, check the margin for a link to the appropriate section (in this example, [sect. 1.4](#)).

[section 1.4](#)

1.3 The future as in a mirror

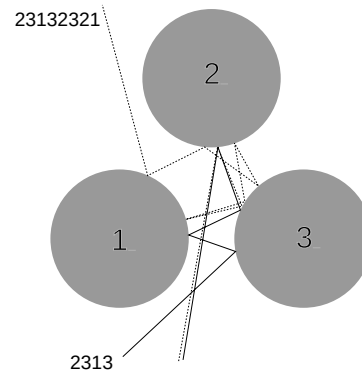
All you need to know about chaos is contained in the introduction of [*ChaosBook*]. However, in order to understand the introduction you will first have to read the rest of the book.

—Gary Morriss

That deterministic dynamics leads to chaos is no surprise to anyone who has tried pool, billiards or snooker—the game is about beating chaos—so we start our story about what chaos is, and what to do about it, with a game of *pinball*. This might seem a trifle, but the game of pinball is to chaotic dynamics what a pendulum is to integrable systems: thinking clearly about what ‘chaos’ in a game of pinball is will help us tackle more difficult problems, such as computing the diffusion constant of a deterministic gas, the drag coefficient of a turbulent boundary layer, or the helium spectrum.

We all have an intuitive feeling for what a ball does as it bounces among the pinball machine's disks, and only high-school level Euclidean geometry is needed to describe its trajectory. A physicist's pinball game is the game of pinball stripped to its bare essentials: three equidistantly placed reflecting disks in a plane, [figure 1.1](#). A physicist's pinball is free, frictionless, point-like, spin-less, perfectly elastic, and noiseless. Point-like pinballs are shot at the disks from random starting

Figure 1.2: Sensitivity to initial conditions: two pin-balls that start out very close to each other separate exponentially with time.



positions and angles; they spend some time bouncing between the disks and then escape.

At the beginning of the 18th century Baron Gottfried Wilhelm Leibniz was confident that given the initial conditions one knew everything a deterministic system would do far into the future. He wrote [25], anticipating by a century and a half the oft-quoted Laplace’s “Given for one instant an intelligence which could comprehend all the forces by which nature is animated...”:

That everything is brought forth through an established destiny is just as certain as that three times three is nine. [...] If, for example, one sphere meets another sphere in free space and if their sizes and their paths and directions before collision are known, we can then foretell and calculate how they will rebound and what course they will take after the impact. Very simple laws are followed which also apply, no matter how many spheres are taken or whether objects are taken other than spheres. From this one sees then that everything proceeds mathematically—that is, infallibly—in the whole wide world, so that if someone could have a sufficient insight into the inner parts of things, and in addition had remembrance and intelligence enough to consider all the circumstances and to take them into account, he would be a prophet and would see the future in the present as in a mirror.

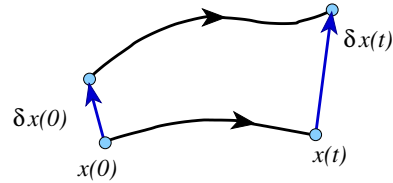
Leibniz chose to illustrate his faith in determinism precisely with the type of physical system that we shall use here as a paradigm of ‘chaos.’ His claim is wrong in a deep and subtle way: a state of a physical system can *never* be specified to infinite precision, and by this we do not mean that eventually the Heisenberg uncertainty principle kicks in. In deterministic dynamics there is no way to take all the circumstances into account, and a single trajectory cannot be tracked, only a ball of nearby initial points makes physical sense.

1.3.1 What is ‘chaos’?

I accept chaos. I am not sure that it accepts me.

—Bob Dylan, *Bringing It All Back Home*

A deterministic system is a system whose present state is *in principle* fully determined by its initial conditions.

Figure 1.3: Unstable trajectories separate with time.

In contrast, radioactive decay, Brownian motion and heat flow are examples of stochastic systems, for which the initial conditions determine the future only partially, due to noise, or other external circumstances beyond our control: the present state reflects the past initial conditions plus the particular realization of the noise encountered along the way.

A deterministic system with sufficiently complicated dynamics can appear to us to be stochastic; disentangling the deterministic from the stochastic is the main challenge in many real-life settings, from stock markets to palpitations of chicken hearts. So, what is ‘chaos’?

In a game of pinball, any two trajectories that start out very close to each other separate exponentially with time. During a finite (and in practice, a very small) number of bounces, the separation $\delta \mathbf{x}(t)$ of these trajectories attains the magnitude L , which is the characteristic linear extent of the whole system (see figure 1.2). This property of *sensitivity to initial conditions* can be quantified as

$$|\delta \mathbf{x}(t)| \approx e^{\lambda t} |\delta \mathbf{x}(0)|$$

where λ , the mean rate of separation of trajectories of the system, is called the *Lyapunov exponent*. For any finite accuracy $\delta x = |\delta \mathbf{x}(0)|$ of the initial data, the dynamics is predictable only up to a finite *Lyapunov time*

chapter 6

$$T_{\text{Lyap}} \approx -\frac{1}{\lambda} \ln |\delta x/L|, \quad (1.1)$$

despite the deterministic and, for Baron Leibniz, infallible simple laws that rule the pinball motion.

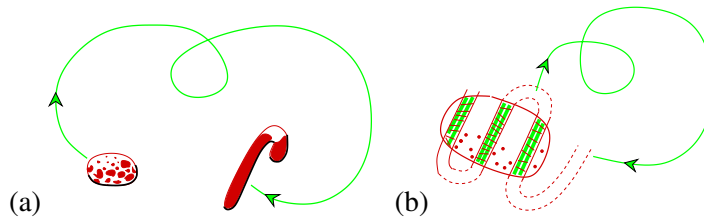
A positive Lyapunov exponent does not in itself lead to chaos. One could try to play 1- or 2-disk pinball game, but it would not be much of a game; trajectories would only separate, never to meet again. What is also needed is *mixing*, the coming together again and again of trajectories. While locally the nearby trajectories separate, the interesting dynamics is confined to a globally finite region of the state space and thus the separated trajectories are necessarily folded back and can re-approach each other arbitrarily closely, infinitely many times. For the case at hand there are 2^n topologically distinct n bounce trajectories that originate from a given disk. More generally, the number of distinct trajectories with n bounces can be quantified as

section 18.1

$$N(n) \approx e^{hn}$$

where h , the growth rate of the number of topologically distinct trajectories, is called the “*topological entropy*” ($h = \ln 2$ in the case at hand).

Figure 1.4: Dynamics of a *chaotic* dynamical system is (a) everywhere locally unstable (positive Lyapunov exponent) and (b) globally mixing (positive entropy). (A. Johansen)



The appellation ‘chaos’ is a confusing misnomer. In deterministic dynamics, there is no chaos in the everyday sense of the word; everything proceeds mathematically—that is, as Baron Leibniz would have it, infallibly. When a physicist says that a certain system exhibits ‘chaos’, she or he means that the system obeys deterministic laws of evolution, but that the outcome is highly sensitive to small uncertainties in the initial state. The word ‘chaos’ has in this context taken on a narrow technical meaning. If a deterministic system is unstable locally (positive Lyapunov exponent) and exhibits mixing globally (positive entropy), it is said to be *chaotic*, figure 1.4. Everybody seems to get their name affixed to yet another definition of ‘chaos’, so here come our, ChaosBook definition:

You can run, but you can’t hide.

While mathematically correct, the definition of chaos as ‘positive Lyapunov and positive entropy’ is useless in practice. Furthermore, a measurement of these quantities is intrinsically asymptotic and beyond reach for natural systems. More powerful is Poincaré’s vision of chaos as the interplay of local instability (unstable periodic orbits) and global mixing (intertwining of their stable and unstable manifolds). In a chaotic system any open ball of initial conditions, no matter how small, will in a finite time overlap with any other finite region and in this sense spread entirely over the asymptotically accessible state space. Once this is grasped, the focus of theory shifts from attempting to predict individual trajectories (which is impossible) to describing the space of possible outcomes and evaluating averages over this space. How this is accomplished is what ChaosBook is about.

1.3.2 What is ‘turbulence’?

I know it when I see it.

—Justice Potter Stewart, *Jacobellis v. Ohio* (1964)

A definition of ‘turbulence’ is even harder to come by. Can you recognize turbulence when you see it? The word comes from ‘tourbillon’, French for ‘vortex’, and intuitively it refers to irregular behavior of spatially extended systems described by deterministic equations of motion—say, a bucket of sloshing water described by the Navier-Stokes equations. But in practice the word ‘turbulence’ tends to refer to messy dynamics which we understand poorly. As soon as a phenomenon is understood better, it is reclaimed and renamed: ‘a route to chaos’, ‘spatiotemporal chaos’, and so on.

chapter 30

Even a baby nonlinear problem can bedevil the smoothest dynamicist, and thus there is much squabbling about naming different kinds of complex dynamics exhibited by nonlinear flows. In practice, “chaos” tends to refer to unstable $3d$ flows ($1d$ and $2d$ maps). If the dimension is higher, new names are made up. For example, if most orbits of a system are unstable to perturbations in two real eigendirections, that is “hyperchaos.” A waste of a hyperbole that could have been saved up to describe a phenomenon of a greater generality than the number 2.

Flows described by partial differential equations [PDEs] are said to be infinite dimensional, because infinitely many ordinary differential equations [ODEs] are needed to represent the dynamics of one PDE. Even though their state space is ‘infinite-dimensional’, the long-time dynamics of viscous flows, such as Navier-Stokes, and PDEs modeling them, such as Kuramoto-Sivashinsky, exhibits, when dissipation is high and the system spatial extent small, apparent ‘low-dimensional’ dynamical behaviors. For some of these the asymptotic dynamics is known to be confined to a finite-dimensional *inertial manifold*, though the rigorous upper bounds on this dimension are not of much use in the practice.

chapter 30

For large spatial extent the complexity of the spatial motions also needs to be taken into account. The systems whose spatial correlations decay sufficiently fast, and the attractor dimension and number of positive Lyapunov exponents diverges with system size are said to be extensively, ‘spatio-temporally chaotic’ or ‘weakly turbulent.’ Spatio-temporally chaotic systems are characterized by creation / annihilation of ‘defects.’ They are extensive; if you increase the spatial extent in a given direction by a factor of two, you will need twice as many ‘computational degrees of freedom’ to describe it to the same accuracy. Conversely, for small system sizes the accurate description might require a large set of coupled ODEs, but dynamics can still be ‘low-dimensional’ in the sense that it is characterized by one or a few positive Lyapunov exponents. There is no wide range of scales involved, nor decays of spatial correlations, and the system is in this sense only ‘chaotic.’

For a subset of physicists and mathematicians who study idealized ‘fully developed’, ‘homogenous’ turbulence the generally accepted usage is that the ‘turbulent’ fluid is characterized by a range of scales and energy or enstrophy cascades describable by statistic assumptions. What experimentalists, engineers, geophysicists, astrophysicists actually observe looks nothing like a ‘fully developed turbulence.’ In the physically driven wall-bounded shear flows, the turbulence is dominated by unstable *coherent structures*, that is, localized recurrent vortices, rolls, streaks and like. The statistical assumptions fail, and a dynamical systems description from first principles is called for.

appendix A1.5

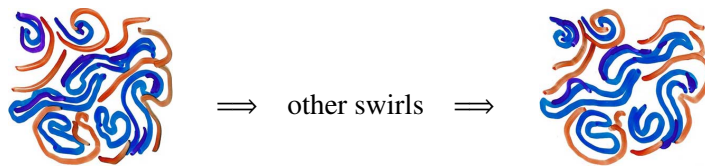
Here comes our quandary. If we ban the words ‘turbulence’ and ‘spatiotemporal chaos’ from our study of small extent systems, the relevance of what we do to larger systems is obscured. The exact unstable coherent structures we determine pertain not only to the spatially small ‘chaotic’ systems, but also the spatially large ‘spatiotemporally chaotic’ and the spatially very large ‘turbulent’ systems. The key aspect we study here - continuous spatial symmetry of the system - is pertinent to all these systems, independent of their size. So, for the lack of more

precise nomenclature, we take the liberty of using the terms ‘chaos’, ‘spatiotemporal chaos’, and ‘turbulence’ interchangeably.

remark 30.1

We return to these painful questions in chapter 30.

In ChaosBook we shall develop a theory of chaotic dynamics for low dimensional attractors visualized as a succession of nearly periodic but unstable motions. In the same spirit, we shall think of turbulence in spatially extended systems in terms of recurrent spatiotemporal patterns. Pictorially, dynamics drives a given spatially extended system (clouds, say) through a repertoire of unstable patterns; as we watch a turbulent system evolve, every so often we catch a glimpse of a familiar pattern:



For any finite spatial resolution, a deterministic flow follows approximately for a finite time an unstable pattern belonging to a finite alphabet of admissible patterns, and the long term dynamics can be thought of as a walk through the space of such patterns. In ChaosBook we recast this image into mathematics.

1.3.3 When does ‘chaos’ matter?

In dismissing Pollock’s fractals because of their limited magnification range, Jones-Smith and Mathur would also dismiss half the published investigations of physical fractals.

— Richard P. Taylor [20, 46]

When should we be mindful of chaos? The solar system is ‘chaotic’, yet we have no trouble keeping track of the annual motions of planets. The rule of thumb is this; if the Lyapunov time (1.1)—the time by which a state space region initially comparable in size to the observational accuracy extends across the entire accessible state space—is significantly shorter than the observational time, you need to master the theory that will be developed here. That is why the main successes of the theory are in statistical mechanics, quantum mechanics, and questions of long term stability in celestial mechanics.

In science popularizations too much has been made of the impact of ‘chaos theory’, so a number of caveats are already needed at this point.

At present the theory that will be developed here is in practice applicable only to systems of a low intrinsic *dimension* – the minimum number of coordinates necessary to capture its essential dynamics. If the system is very turbulent (a description of its long time dynamics requires a space of high intrinsic dimension) we are

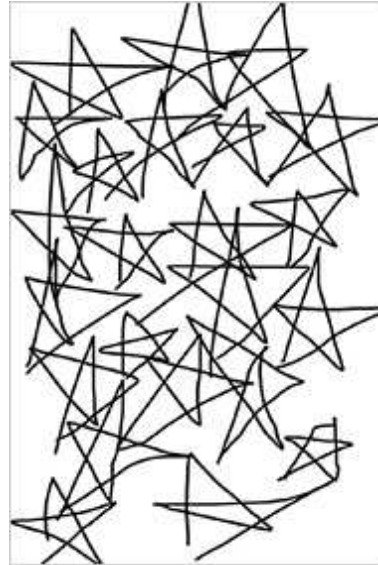


Figure 1.5: Katherine Jones-Smith, 'Untitled 5', the drawing used by K. Jones-Smith and H. Mathur to test the fractal analysis of Pollock's drip paintings [22].

out of luck. Hence insights that the theory offers in elucidating problems of fully developed turbulence, quantum field theory of strong interactions and early cosmology have been modest at best. Even that is a caveat with qualifications. There are applications—such as spatially extended (non-equilibrium) systems, plumber's turbulent pipes, etc.,—where the few important degrees of freedom can be isolated and studied profitably by methods to be described here.

Thus far the theory has had limited practical success when applied to the very noisy systems so important in the life sciences and in economics. Even though we are often interested in phenomena taking place on time scales much longer than the intrinsic time scale (neuronal inter-burst intervals, cardiac pulses, etc.), disentangling 'chaotic' motions from the environmental noise has been very hard.

In 1980's something happened that might be without parallel; this is an area of science where the advent of cheap computation had actually subtracted from our collective understanding. The computer pictures and numerical plots of fractal science of the 1980's have overshadowed the deep insights of the 1970's, and these pictures have since migrated into textbooks. By a regrettable oversight, ChaosBook has none, so 'Untitled 5' of figure 1.5 will have to do as the illustration of the power of fractal analysis. Fractal science posits that certain quantities (Lyapunov exponents, generalized dimensions, ...) can be estimated on a computer. While some of the numbers so obtained are indeed mathematically sensible characterizations of fractals, they are in no sense observable and measurable on the length-scales and time-scales dominated by chaotic dynamics.

remark 1.8

1.4 A game of pinball

Formulas hamper the understanding.
—S. Smale

We are now going to get down to the brass tacks. Time to fasten your seat belts and turn off all electronic devices. But first, a disclaimer: If you understand the rest of this chapter on the first reading, you either do not need this book, or you are delusional. If you do not understand it, it is not because the people who figured all this out first are smarter than you: the most you can hope for at this stage is to get a flavor of what lies ahead. If a statement in this chapter mystifies/intrigues, fast forward to a section indicated by [\[section ...\]](#) on the margin, read only the parts that you feel you need. Of course, we think that you need to learn ALL of it, or otherwise we would not have included it in ChaosBook in the first place.

Confronted with a potentially chaotic dynamical system, our analysis proceeds in three stages; I. diagnose, II. count, III. measure. First, we determine the intrinsic *dimension* of the system—the minimum number of coordinates necessary to capture its essential dynamics. If the system is very turbulent we are, at present, out of luck. We know only how to deal with the transitional regime between regular motions and chaotic dynamics in a few dimensions. That is still something; even an infinite-dimensional system such as a burning flame front can turn out to have a very few chaotic degrees of freedom. In this regime the chaotic dynamics is restricted to a space of low dimension, the number of relevant parameters is small, and we can proceed to step II; we *count* and *classify* all possible topologically distinct trajectories of the system into a hierarchy whose successive layers require increased precision and patience on the part of the observer. This we shall do in sect. [1.4.2](#). If successful, we can proceed with step III: investigate the *weights* of the different pieces of the system.

chapter 14

chapter 18

We commence our analysis of the pinball game with steps I, II: diagnose, count. We shall return to step III—measure—in sect. [1.5](#). The three sections that follow are *highly* technical, they go into the guts of what the book is about. If today is not your thinking day, skip them, jump straight to sect. [1.7](#).

chapter 23

1.4.1 Symbolic dynamics

With the game of pinball we are in luck—it is a low dimensional system, free motion in a plane. The motion of a point particle is such that after a collision with one disk it either continues to another disk or it escapes. If we label the three disks by 1, 2 and 3, we can associate every trajectory with an *itinerary*, a sequence of labels indicating the order in which the disks are visited; for example, the two trajectories in figure [1.2](#) have itineraries [_2313_](#), [_23132321_](#) respectively. Such labeling goes by the name *symbolic dynamics*. As the particle cannot collide two times in succession with the same disk, any two consecutive symbols must differ. This is an example of *pruning*, a rule that forbids certain subsequences of symbols. Deriving pruning rules is in general a difficult problem, but with the game of pinball we are lucky—for well-separated disks there are no further pruning rules.

exercise 1.1

section 2.1

chapter 15

The choice of symbols is in no sense unique. For example, as at each bounce we can either proceed to the next disk or return to the previous disk, the above 3-letter alphabet can be replaced by a binary $\{0, 1\}$ alphabet, figure [1.6](#). A clever

Figure 1.6: Binary labeling of the 3-disk pinball trajectories; a bounce in which the trajectory returns to the preceding disk is labeled 0, and a bounce which results in continuation to the third disk is labeled 1.

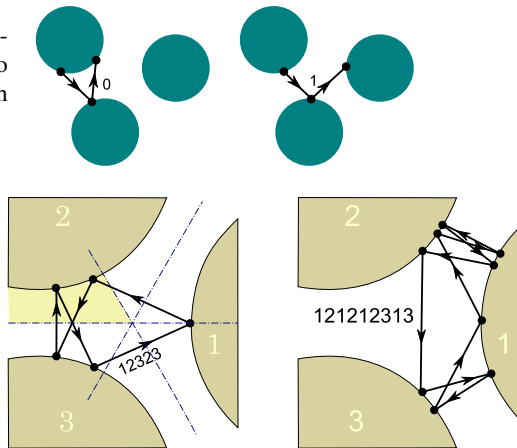


Figure 1.7: The 3-disk pinball cycles $\overline{12323}$ and $\overline{121212313}$.

choice of an alphabet will incorporate important features of the dynamics, such as its symmetries.

section 14.6

Suppose you wanted to play a good game of pinball, that is, get the pinball to bounce as many times as you possibly can—what would be a winning strategy? The simplest thing would be to try to aim the pinball so it bounces many times between a pair of disks—if you managed to shoot it so it starts out in the periodic orbit bouncing along the line connecting two disk centers, it would stay there forever. Your game would be just as good if you managed to get it to keep bouncing between the three disks forever, or place it on any periodic orbit. The only rub is that any such orbit is *unstable*, so you have to aim very accurately in order to stay close to it for a while. So it is pretty clear that if one is interested in playing well, unstable periodic orbits are important—they form the *skeleton* onto which all trajectories trapped for long times cling.

1.4.2 Partitioning with periodic orbits

A trajectory is periodic if it returns to its starting position and momentum. We shall sometimes refer to the set of periodic points that belong to a given periodic orbit as a *cycle*.

Short periodic orbits are easily drawn and enumerated—an example is drawn in figure 1.7—but it is rather hard to perceive the systematics of orbits from their configuration space shapes. In mechanics a trajectory is fully and uniquely specified by its position and momentum at a given instant, and no two distinct state space trajectories can intersect. Their projections onto arbitrary subspaces, however, can and do intersect, in rather unilluminating ways. In the pinball example the problem is that we are looking at the projections of a 4-dimensional state space trajectories onto a 2-dimensional subspace, the configuration space. A clearer picture of the dynamics is obtained by constructing a set of state space Poincaré sections.

Suppose that the pinball has just bounced off disk 1. Depending on its position and outgoing angle, it could proceed to either disk 2 or 3. Not much happens in

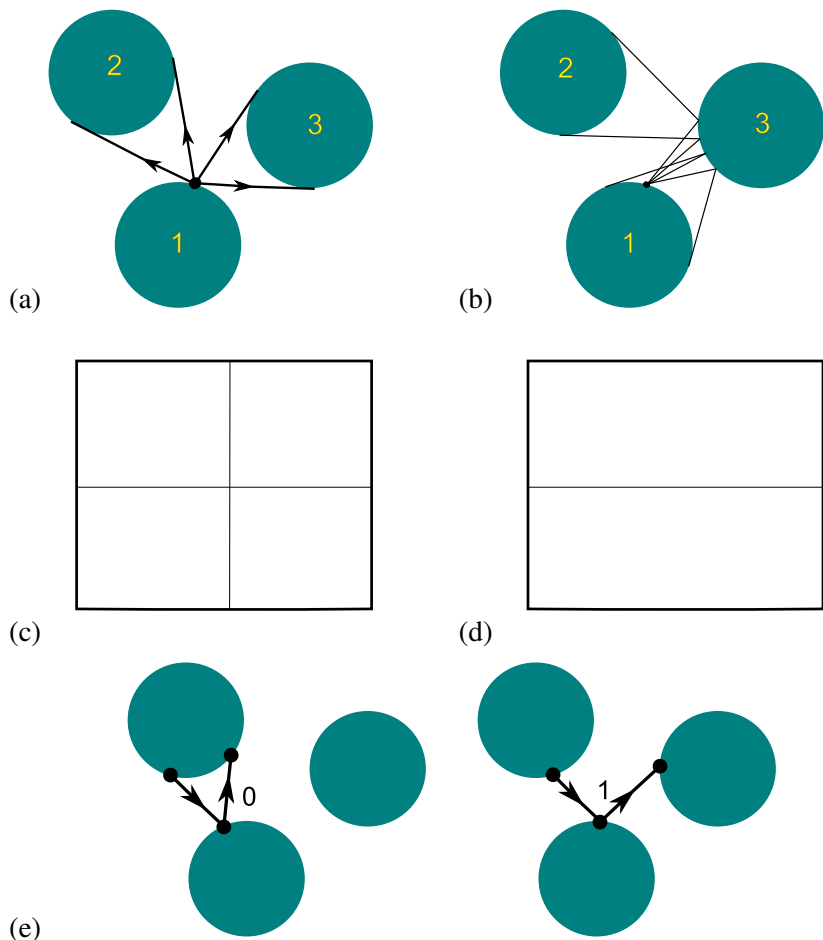


Figure 1.8: Starting out from figure 9.2 (a) initial point (s_0, p_0) , a pinball can (a) either hit another disk, escape, or (b) hit two disks in a row. Trajectories emanating from the disk 1 form phase-space strips (c) $\mathcal{M}_{12}$, $\mathcal{M}_{13}$ that reach either disk 2, or disk 3 in one bounce, or (d) $\mathcal{M}_{121}$, $\mathcal{M}_{131}$, $\mathcal{M}_{132}$ and $\mathcal{M}_{123}$ that reach a disk in two bounces. Hitting disks in a row requires a sharper aim, with initial conditions that hit consecutive disks nested within each other. (e) Binary labeling of equilateral 3-disk pinball trajectories: a bounce in which the trajectory returns to the preceding disk is labeled 0, and a bounce which results in continuation to the third disk is labeled 1. (Phase space strips plotted for disk radius : center separation ratio $a:R = 1:2.5$).

between the bounces—the ball just travels at constant velocity along a straight line—so we can reduce the 4-dimensional flow to a 2-dimensional map P that takes the coordinates of the pinball from one disk edge to another disk edge. The trajectory just after the moment of impact is defined by s_n , the arc-length position of the n th bounce along the billiard wall, and $p_n = p \sin \phi_n$ the momentum component parallel to the billiard wall at the point of impact. Such section of a flow is called a *Poincaré section*. In terms of Poincaré sections, the dynamics is reduced to the set of six maps $P_{s_k \leftarrow s_j} : (s_n, p_n) \mapsto (s_{n+1}, p_{n+1})$, with $s \in \{1, 2, 3\}$, from the boundary of the disk j to the boundary of the next disk k .

example 15.8

chapter 9

Next, we mark in the Poincaré section those initial conditions which do not escape in one bounce. There are two strips of survivors, as the trajectories originating from one disk can hit either of the other two disks, or escape without further ado. We label the two strips $\mathcal{M}_{12}$, $\mathcal{M}_{13}$. Embedded within them there are four strips $\mathcal{M}_{121}$, $\mathcal{M}_{123}$, $\mathcal{M}_{131}$, $\mathcal{M}_{132}$ of initial conditions that survive for two bounces, and so forth. see figure 1.8. Provided that the disks are sufficiently separated, after n bounces the survivors are divided into 2^n distinct strips: the $\mathcal{M}_i$ th strip consists of all points with itinerary $i = s_1 s_2 s_3 \dots s_n$, $s = \{1, 2, 3\}$. The unstable cycles as a skeleton of chaos are almost visible here: each such patch contains a periodic point $\overline{s_1 s_2 s_3 \dots s_n}$ with the basic block infinitely repeated. Pe-

riodic points are skeletal in the sense that as we look further and further, the strips shrink but the periodic points stay put forever.

We see now why it pays to utilize a symbolic dynamics; it provides a navigation chart through chaotic state space. There exists a unique trajectory for every admissible infinite length itinerary, and a unique itinerary labels every trapped trajectory. For example, the only trajectory labeled by $\overline{12}$ is the 2-cycle bouncing along the line connecting the centers of disks 1 and 2; any other trajectory starting out as $12\dots$ either eventually escapes or hits the 3rd disk.

1.4.3 Escape rate

example 20.4

What is a good physical quantity to compute for the game of pinball? Such a system, for which almost any trajectory eventually leaves a finite region (the pinball table) never to return, is said to be open, or a *repeller*. The repeller *escape rate* is an eminently measurable quantity. An example of such a measurement would be an unstable molecular or nuclear state which can be well approximated by a potential with the possibility of escape in certain directions. In an experiment many projectiles are injected into a macroscopic ‘black box’ enclosing a microscopic non-confining short-range potential, and their mean escape rate is measured, as in figure 1.1. The numerical experiment might consist of injecting the pinball between the disks in some random direction and asking how many times the pinball bounces on the average before it escapes the region between the disks.

exercise 1.2

For a theorist, a good game of pinball consists in predicting accurately the asymptotic lifetime (or the escape rate) of the pinball. We now show how periodic orbit theory accomplishes this for us. Each step will be so simple that you can follow even at the cursory pace of this overview, and still the result is surprisingly elegant.

Consider figure 1.8 again. In each bounce the initial conditions get thinned out, yielding twice as many thin strips as at the previous bounce. The total area that remains at a given time is the sum of the areas of the strips, so that the fraction of survivors after n bounces, or the *survival probability* is given by

$$\begin{aligned}\hat{\Gamma}_1 &= \frac{|\mathcal{M}_0|}{|\mathcal{M}|} + \frac{|\mathcal{M}_1|}{|\mathcal{M}|}, & \hat{\Gamma}_2 &= \frac{|\mathcal{M}_{00}|}{|\mathcal{M}|} + \frac{|\mathcal{M}_{10}|}{|\mathcal{M}|} + \frac{|\mathcal{M}_{01}|}{|\mathcal{M}|} + \frac{|\mathcal{M}_{11}|}{|\mathcal{M}|}, \\ \hat{\Gamma}_n &= \frac{1}{|\mathcal{M}|} \sum_i^{(n)} |\mathcal{M}_i|,\end{aligned}\tag{1.2}$$

where i is a label of the i th strip, $|\mathcal{M}|$ is the initial area, and $|\mathcal{M}_i|$ is the area of the i th strip of survivors. $i = 01, 10, 11, \dots$ is a label, not a binary number. Since at each bounce one routinely loses about the same fraction of trajectories, one expects the sum (1.2) to fall off exponentially with n and tend to the limit

chapter 27

$$\hat{\Gamma}_{n+1}/\hat{\Gamma}_n = e^{-\gamma_n} \rightarrow e^{-\gamma}.\tag{1.3}$$

The quantity γ is called the *escape rate* from the repeller.

1.5 Chaos for cyclists

Étant données des équations ... et une solution particulière quelconque de ces équations, on peut toujours trouver une solution périodique (dont la période peut, il est vrai, être très longue), telle que la différence entre les deux solutions soit aussi petite qu'on le veut, pendant un temps aussi long qu'on le veut. D'ailleurs, ce qui nous rend ces solutions périodiques si précieuses, c'est qu'elles sont, pour ainsi dire, la seule brèche par où nous puissions essayer de pénétrer dans une place jusqu'ici réputée inabordable.

—H. Poincaré, *Les méthodes nouvelles de la mécanique céleste*

We shall now show that the escape rate γ can be extracted from a highly convergent *exact* expansion by reformulating the sum (1.2) in terms of unstable periodic orbits.

If, when asked what the 3-disk escape rate is for a disk of radius 1, center-center separation 6, velocity 1, you answer that the continuous time escape rate is roughly $\gamma = 0.4103384077693464893384613078192 \dots$, you do not need this book. If you have no clue, hang on.

1.5.1 How big is my neighborhood?

Of course, we can prove all these results directly from Eq. (20.15) by pedestrian mathematical manipulations, but that only makes it harder to appreciate their physical significance.

—Rick Salmon, “Lectures on Geophysical Fluid Dynamics”, Oxford Univ. Press (1998)

Not only do the periodic points keep track of topological ordering of the strips, but, as we shall now show, they also determine their size. As a trajectory evolves, it carries along and distorts its infinitesimal neighborhood. Let

$$x(t) = f^t(x_0)$$

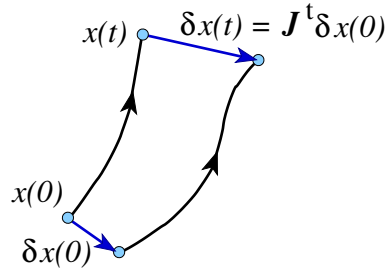
denote the trajectory of an initial point $x_0 = x(0)$. Expanding $f^t(x_0 + \delta x_0)$ to linear order, the evolution of the distance to a neighboring trajectory $x(t) + \delta x(t)$ is given by the Jacobian matrix J :

$$\delta x_i(t) = \sum_{j=1}^d J^t(x_0)_{ij} \delta x_{0j}, \quad J^t(x_0)_{ij} = \frac{\partial x_i(t)}{\partial x_{0j}}. \quad (1.4)$$

A trajectory of a pinball moving on a flat surface is specified by two position coordinates and the direction of motion, so in this case $d = 3$. Evaluation of a cycle Jacobian matrix is a long exercise - here we just state the result. The Jacobian matrix describes the deformation of an infinitesimal neighborhood of

section 9.2

Figure 1.9: The Jacobian matrix J^t maps an infinitesimal displacement δx at x_0 into a displacement $J^t(x_0)\delta x$ a finite time t later.



$x(t)$ along the flow; its eigenvectors and eigenvalues give the directions and the corresponding rates of expansion or contraction, figure 1.9. The trajectories that start out in an infinitesimal neighborhood separate along the unstable directions (those whose eigenvalues are greater than unity in magnitude), approach each other along the stable directions (those whose eigenvalues are less than unity in magnitude), and change their distance only sub-exponentially (or not at all) along the marginal directions (those whose eigenvalues equal unity in magnitude).

In our game of pinball the beam of neighboring trajectories is defocused along the unstable eigen-direction of the Jacobian matrix J .

As the heights of the strips in figure 1.8 are effectively constant, we can concentrate on their thickness. If the height is $\approx L$, then the area of the i th strip is $\mathcal{M}_i \approx Ll_i$ for a strip of width l_i .

Each strip i in figure 1.8 contains a periodic point x_i . The finer the intervals, the smaller the variation in flow across them, so the contribution from the strip of width l_i is well-approximated by the contraction around the periodic point x_i within the interval,

$$l_i = a_i / |\Lambda_i|, \quad (1.5)$$

where Λ_i is the unstable eigenvalue of the Jacobian matrix $J^t(x_i)$ evaluated at the i th periodic point for $t = T_p$, the full period (due to the low dimensionality, the Jacobian can have at most one unstable eigenvalue). Only the magnitude of this eigenvalue matters, we can disregard its sign. The prefactors a_i reflect the overall size of the system and the particular distribution of starting values of x . As the asymptotic trajectories are strongly mixed by bouncing chaotically around the repeller, we expect their distribution to be insensitive to smooth variations in the distribution of initial points.

section 19.4

To proceed with the derivation we need the *hyperbolicity* assumption: for large n the prefactors $a_i \approx O(1)$ are overwhelmed by the exponential growth of Λ_i , so we neglect them. If the hyperbolicity assumption is justified, we can replace $|\mathcal{M}_i| \approx Ll_i$ in (1.2) by $1/|\Lambda_i|$ and consider the sum

section 21.1.1

$$\Gamma_n = \sum_i^{(n)} 1/|\Lambda_i|,$$

where the sum goes over all periodic points of period n . We now define a generating function for sums over all periodic orbits of all periods:

remark 18.1

$$\Gamma(z) = \sum_{n=1}^{\infty} \Gamma_n z^n. \quad (1.6)$$

Recall that for large n the n th level sum (1.2) tends to the limit $\Gamma_n \rightarrow e^{-n\gamma}$, so the escape rate γ is determined by the smallest $z = e^\gamma$ for which (1.6) diverges:

$$\Gamma(z) \approx \sum_{n=1}^{\infty} (ze^{-\gamma})^n = \frac{ze^{-\gamma}}{1 - ze^{-\gamma}}. \quad (1.7)$$

This is the property of $\Gamma(z)$ that motivated its definition. Next, we devise a formula for (1.6) expressing the escape rate in terms of periodic orbits:

$$\begin{aligned} \Gamma(z) &= \sum_{n=1}^{\infty} z^n \sum_i^{(n)} |\Lambda_i|^{-1} \\ &= \frac{z}{|\Lambda_0|} + \frac{z}{|\Lambda_1|} + \frac{z^2}{|\Lambda_{00}|} + \frac{z^2}{|\Lambda_{01}|} + \frac{z^2}{|\Lambda_{10}|} + \frac{z^2}{|\Lambda_{11}|} \\ &\quad + \frac{z^3}{|\Lambda_{000}|} + \frac{z^3}{|\Lambda_{001}|} + \frac{z^3}{|\Lambda_{010}|} + \frac{z^3}{|\Lambda_{100}|} + \dots \end{aligned} \quad (1.8)$$

For sufficiently small z this sum is convergent. The escape rate γ is now given by the leading pole of (1.7), rather than by a numerical extrapolation of a sequence of γ_n extracted from (1.3). As any finite truncation $n < n_{\text{trunc}}$ of (1.8) is a polynomial in z , convergent for any z , finding this pole requires that we know something about Γ_n for any n , and that might be a tall order.

section 21.3

We could now proceed to estimate the location of the leading singularity of $\Gamma(z)$ from finite truncations of (1.8) by methods such as Padé approximants. However, as we shall now show, it pays to first perform a simple resummation that converts this divergence into a *zero* of a related function.

1.5.2 Dynamical zeta function

If a trajectory retraces a *prime* cycle r times, its expanding eigenvalue is Λ_p^r . A prime cycle p is a single traversal of the orbit; its label is a non-repeating symbol string of n_p symbols. There is only one prime cycle for each cyclic permutation class. For example, $p = 0011 = 1001 = 1100 = 0110$ is prime, but $0101 = 01$ is not.

By the chain rule for derivatives the stability of a cycle is the same everywhere along the orbit, so each prime cycle of period n_p contributes n_p terms to the sum (1.8). Hence (1.8) can be rewritten as

exercise 18.2
section 4.5

$$\Gamma(z) = \sum_p n_p \sum_{r=1}^{\infty} \left(\frac{z^{n_p}}{|\Lambda_p|} \right)^r = \sum_p \frac{n_p t_p}{1 - t_p}, \quad t_p = \frac{z^{n_p}}{|\Lambda_p|} \quad (1.9)$$

where the index p runs through all distinct *prime* cycles. Note that we have resummed the contribution of the cycle p to all times, so truncating the summation up to given p is *not* a finite time $n \leq n_p$ approximation, but an asymptotic, *infinite*

time estimate based by approximating stabilities of all cycles by a finite number of the shortest cycles and their repeats. The $n_p z^{n_p}$ factors in (1.9) suggest rewriting the sum as a derivative

$$\Gamma(z) = -z \frac{d}{dz} \sum_p \ln(1 - t_p).$$

Hence $\Gamma(z)$ is a logarithmic derivative derivative of the infinite product

$$1/\zeta(z) = \prod_p (1 - t_p), \quad t_p = \frac{z^{n_p}}{|\Lambda_p|}. \quad (1.10)$$

This function is called the *dynamical zeta function*, in analogy to the Riemann zeta function, which motivates the ‘zeta’ in its definition as $1/\zeta(z)$. This is the prototype formula of periodic orbit theory. The zero of $1/\zeta(z)$ is a pole of $\Gamma(z)$, and the problem of estimating the asymptotic escape rates from finite n sums such as (1.2) is now reduced to a study of the zeros of the dynamical zeta function (1.10). The escape rate is related by (1.7) to a divergence of $\Gamma(z)$, and $\Gamma(z)$ diverges whenever $1/\zeta(z)$ has a zero.

section 27.1

section 22.4

Easy, you say: “Zeros of (1.10) can be read off the formula, a zero

$$z_p = |\Lambda_p|^{1/n_p}$$

for each term in the product. What’s the problem?” Dead wrong!

1.5.3 Cycle expansions

How are formulas such as (1.10) used? We start by computing the periods and eigenvalues of the shortest cycles. This usually requires some numerical work, such as the Newton method searches for periodic solutions; we shall assume that the numerics are under control, and that *all* short cycles up to given period have been found. In our pinball example this can be done by elementary geometrical optics. It is very important not to miss any short cycles, as the calculation is as accurate as the shortest cycle dropped—including cycles longer than the shortest omitted does not improve the accuracy. The result of such numerics is a table of the shortest cycles, their periods and their stabilities.

chapter 16

section 34.3

Now expand the infinite product (1.10), grouping together the terms of the same total symbol string length

$$\begin{aligned} 1/\zeta &= (1 - t_0)(1 - t_1)(1 - t_{10})(1 - t_{100}) \cdots \\ &= 1 - t_0 - t_1 - [t_{10} - t_1 t_0] - [(t_{100} - t_{10} t_0) + (t_{101} - t_{10} t_1)] \\ &\quad - [(t_{1000} - t_0 t_{100}) + (t_{1110} - t_1 t_{110}) \\ &\quad + (t_{1001} - t_1 t_{001} - t_{101} t_0 + t_{10} t_0 t_1)] - \dots \end{aligned} \quad (1.11)$$

The virtue of the expansion is that the sum of all terms of the same total period

chapter 23

Figure 1.10: Approximation to a smooth dynamics (left frame) by the skeleton of periodic points, together with their linearized neighborhoods, (right frame). Indicated are segments of two 1-cycles and a 2-cycle that alternates between the neighborhoods of the two 1-cycles, shadowing first one of the two 1-cycles, and then the other.

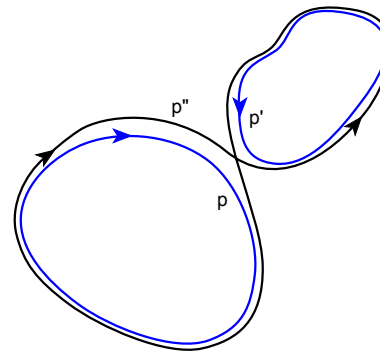
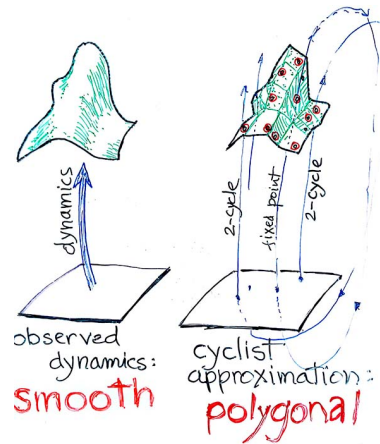


Figure 1.11: A longer cycle p'' shadowed by a pair (a 'pseudo orbit') of shorter cycles p and p' .

n (grouped in brackets above) is a number that is exponentially smaller than a typical term in the sum, for geometrical reasons we explain in the next section.

section 23.1

The calculation is now straightforward. We substitute a finite set of the eigenvalues and periods of the shortest prime cycles into the cycle expansion (1.11), and obtain a polynomial approximation to $1/\zeta$. We then vary z in (1.10) and determine the escape rate γ by finding the smallest $z = e^\gamma$ for which (1.11) vanishes.

1.5.4 Shadowing

When you actually start computing this escape rate, you will find out that the convergence is very impressive: only three input numbers (the two fixed points $\bar{0}$, $\bar{1}$ and the 2-cycle $\bar{10}$) already yield the pinball escape rate to 3-4 significant digits! We have omitted an infinity of unstable cycles; so why does approximating the dynamics by a finite number of the shortest cycle eigenvalues work so well?

section 23.2.2

The convergence of cycle expansions of dynamical zeta functions is a consequence of the smoothness and analyticity of the underlying flow. Intuitively, one can understand the convergence in terms of the geometrical picture sketched in figure 1.10; the key observation is that the long orbits are *shadowed* by sequences of shorter orbits.

A typical term in (1.11) is a difference of a long cycle $\{ab\}$ minus its shadowing

approximation by shorter cycles $\{a\}$ and $\{b\}$ (see figure 1.11),

$$t_{ab} - t_a t_b = t_{ab}(1 - t_a t_b / t_{ab}) = t_{ab} \left(1 - \left| \frac{\Lambda_{ab}}{\Lambda_a \Lambda_b} \right| \right), \quad (1.12)$$

where a and b are symbol sequences of the two shorter cycles. If all orbits are weighted equally ($t_p = z^{n_p}$), such combinations cancel exactly; if orbits of similar symbolic dynamics have similar weights, the weights in such combinations almost cancel.

This can be understood in the context of the pinball game as follows. Consider orbits $\overline{0}$, $\overline{1}$ and $\overline{01}$. The first corresponds to bouncing between any two disks while the second corresponds to bouncing successively around all three, tracing out an equilateral triangle. The cycle $\overline{01}$ starts at one disk, say disk 2. It then bounces from disk 3 back to disk 2 then bounces from disk 1 back to disk 2 and so on, so its itinerary is $\overline{2321}$. In terms of the bounce types shown in figure 1.6, the trajectory is alternating between 0 and 1. The incoming and outgoing angles when it executes these bounces are very close to the corresponding angles for 0 and 1 cycles. Also the distances traversed between bounces are similar so that the 2-cycle expanding eigenvalue Λ_{01} is close in magnitude to the product of the 1-cycle eigenvalues $\Lambda_0 \Lambda_1$.

To understand this on a more general level, try to visualize the partition of a chaotic dynamical system's state space in terms of cycle neighborhoods as a tessellation (a tiling) of the dynamical system, with smooth flow approximated by its periodic orbit skeleton, each 'tile' centered on a periodic point, and the scale of the 'tile' determined by the linearization of the flow around the periodic point, as illustrated by figure 1.10.

The orbits that follow the same symbolic dynamics, such as $\{ab\}$ and a 'pseudo orbit' $\{a\}\{b\}$ (see figure 1.11), lie close to each other in state space; long shadowing pairs have to start out exponentially close to beat the exponential growth in separation with time. If the weights associated with the orbits are multiplicative along the flow (for example, by the chain rule for products of derivatives) and the flow is smooth, the term in parenthesis in (1.12) falls off exponentially with the cycle period, and therefore the curvature expansions are expected to be highly convergent.

chapter 28

1.6 Change in time

MEN are deplorably ignorant with respect to natural things and modern philosophers as though dreaming in the darkness must be aroused and taught the uses of things the dealing with things they must be made to quit the sort of learning that comes only from books and that rests only on vain arguments from probability and upon conjectures.

— William Gilbert, *De Magnete*, 1600

The above derivation of the dynamical zeta function formula for the escape rate has one shortcoming; it estimates the fraction of survivors as a function of the number of pinball bounces, but the physically interesting quantity is the escape rate measured in units of continuous time. For continuous time flows, the escape rate (1.2) is generalized as follows. Define a finite state space region $\mathcal{M}$ such that a trajectory that exits $\mathcal{M}$ never reenters. For example, any pinball that falls off the edge of a pinball table in figure 1.1 is gone forever. Start with a uniform distribution of initial points. The fraction of initial x whose trajectories remain within $\mathcal{M}$ at time t is expected to decay exponentially

$$\Gamma(t) = \frac{\int_{\mathcal{M}} dx dy \delta(y - f^t(x))}{\int_{\mathcal{M}} dx} \rightarrow e^{-\gamma t}.$$

The integral over x starts a trajectory at every $x \in \mathcal{M}$. The integral over y tests whether this trajectory is still in $\mathcal{M}$ at time t . The kernel of this integral

$$\mathcal{L}^t(y, x) = \delta(y - f^t(x)) \quad (1.13)$$

is the Dirac delta function, as for a deterministic flow the initial point x maps into a unique point y at time t . For discrete time, $f^n(x)$ is the n th iterate of the map f . For continuous flows, $f^t(x)$ is the trajectory of the initial point x , and it is appropriate to express the finite time kernel $\mathcal{L}^t$ in terms of $\mathcal{A}$, the generator of infinitesimal time translations

$$\mathcal{L}^t = e^{t\mathcal{A}},$$

very much in the way the quantum evolution is generated by the Hamiltonian H , the generator of infinitesimal time quantum transformations.

section 19.6

As the kernel $\mathcal{L}$ is the key to everything that follows, we shall give it a name, and refer to it and its generalizations as the *evolution operator* for a d -dimensional map or a d -dimensional flow.

The number of periodic points increases exponentially with the cycle period (in the case at hand, as 2^n). As we have already seen, this exponential proliferation of cycles is not as dangerous as it might seem; as a matter of fact, all our computations will be carried out in the $n \rightarrow \infty$ limit. Though a quick look at long-time density of trajectories might reveal it to be complex beyond belief, this distribution is still generated by a simple deterministic law, and with some luck and insight, our labeling of possible motions will reflect this simplicity. If the rule that gets us from one level of the classification hierarchy to the next does not depend strongly on the level, the resulting hierarchy is approximately self-similar. We now turn such approximate self-similarity to our advantage, by turning it into an operation, the action of the evolution operator, whose iteration encodes the self-similarity.

1.6.1 Trace formula

In physics, when we do not understand something, we give it a name.

—Matthias Neubert

Figure 1.12: The trace of an evolution operator is concentrated in tubes around prime cycles, of period T_p and thickness $1/|\Lambda_p|^r$ for the r th repetition of the prime cycle p .

$$\begin{aligned}
 \text{tr } \mathcal{L}^t &= \sum_{\alpha=0}^{\infty} e^{-\alpha t} && \text{MIGHT DIVERGE!} \\
 &= \text{[Diagram: A tube around a prime cycle } p \text{ with period } T_p \text{ and thickness } 1/|\Lambda_p|^r \text{]} \\
 &= \sum_p \int_{V_p} dx \mathcal{L}^t(x, x) = \sum_p T_p \sum_{r=1}^{\infty} \frac{\delta(t - rT_p)}{|\det(\mathbf{1} - M_p^r)|} \\
 &&& \text{prime cycles} \quad \text{repeats} \\
 &&& \text{length } \times \text{thickness of prime cycle contribution} \\
 &= \sum_{\text{primes}} \sum_{\text{repeats}} \text{[Diagram: A tube around a prime cycle } p \text{ with period } T_p \text{ and thickness } 1/|\Lambda_p|^r \text{]}
 \end{aligned}$$

Recasting dynamics in terms of evolution operators changes everything. So far our formulation has been heuristic, but in the evolution operator formalism the escape rate and any other dynamical average are given by exact formulas, extracted from the spectra of evolution operators. The key tools are *trace formulas* and *spectral determinants*.

The trace of an operator is given by the sum of its eigenvalues. The explicit expression (1.13) for $\mathcal{L}^t(x, y)$ enables us to evaluate the trace. Identify y with x and integrate x over the whole state space. The result is an expression for $\text{tr } \mathcal{L}^t$ as a sum over neighborhoods of prime cycles p and their repetitions

section 21.2

$$\text{tr } \mathcal{L}^t = \sum_p T_p \sum_{r=1}^{\infty} \frac{\delta(t - rT_p)}{|\det(\mathbf{1} - M_p^r)|}, \quad (1.14)$$

where T_p is the period of prime cycle p , and the monodromy matrix M_p is the flow-transverse part of Jacobian matrix J (1.4). This formula has a simple geometrical interpretation sketched in figure 1.12. After the r th return to a Poincaré section, the initial tube M_p has been stretched out along the expanding eigen-directions, with the overlap with the initial volume given by $1/|\det(\mathbf{1} - M_p^r)| \rightarrow 1/|\Lambda_p|$, the same weight we obtained heuristically in sect. 1.5.1.

The ‘spiky’ sum (1.14) is disquieting in the way reminiscent of the Poisson resummation formulas of Fourier analysis; the left-hand side is the smooth eigenvalue sum $\text{tr } e^{\mathcal{A}t} = \sum e^{s_{\alpha}t}$, while the right-hand side equals zero everywhere except for the set $t = rT_p$. A Laplace transform smooths the sum over Dirac delta functions in cycle periods and yields the *trace formula* for the eigenspectrum $s_0, s_1, \dots$ of the deterministic evolution operator:

chapter 21

$$\begin{aligned}
 \int_{0+}^{\infty} dt e^{-st} \text{tr } \mathcal{L}^t &= \text{tr } \frac{1}{s - \mathcal{A}} = \\
 \sum_{\alpha=0}^{\infty} \frac{1}{s - s_{\alpha}} &= \sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r(\beta A_p - sT_p)}}{|\det(\mathbf{1} - M_p^r)|}. \quad (1.15)
 \end{aligned}$$

The beauty of trace formulas lies in the fact that everything on the right-hand-side—prime cycles p , their periods T_p and the eigenvalues of M_p —is an invariant property of the flow, independent of any coordinate choice.

1.6.2 Spectral determinant

The eigenvalues of a linear operator are given by the zeros of the appropriate determinant. One way to evaluate determinants is to expand them in terms of traces, using the identities

$$\frac{d}{ds} \ln \det(s - \mathcal{A}) = \text{tr} \frac{d}{ds} \ln(s - \mathcal{A}) = \text{tr} \frac{1}{s - \mathcal{A}}, \quad (1.16)$$

exercise 4.1

and integrating over s . In this way the *spectral determinant* of an evolution operator becomes related to the traces that we have just computed:

chapter 22

$$\det(s - \mathcal{A}) = \exp \left(- \sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{-sT_p r}}{|\det(\mathbf{1} - M_p^r)|} \right). \quad (1.17)$$

The $1/r$ factor is due to the s integration, leading to the replacement $T_p \rightarrow T_p/rT_p$ in the periodic orbit expansion (1.15).

section 22.5

We have now retraced the heuristic derivation of the divergent sum (1.7) and the dynamical zeta function (1.10), but this time with no approximations: formula (1.17) is *exact*. The computation of the zeros of $\det(s - \mathcal{A})$ proceeds very much like the computations of sect. 1.5.3.

1.7 From chaos to statistical mechanics

Under heaven, all is chaos. The situation is excellent!
— Chairman Mao Zedong, a letter to Jiang Qing

The replacement of individual trajectories by evolution operators which propagate densities feels like a bit of voodoo. Nevertheless, something very radical and deeply foundational has taken place. Understanding the distinction between evolution of individual trajectories and the evolution of the densities of trajectories is key to understanding statistical mechanics—this is the conceptual basis of the second law of thermodynamics, and the origin of irreversibility of the arrow of time for deterministic systems with time-reversible equations of motion: reversibility is attainable for distributions whose measure in the space of density functions goes exponentially to zero with time.

Consider a chaotic flow, such as the stirring of red and white paint by some deterministic machine. *If* we were able to track individual trajectories, the fluid would forever remain a striated combination of pure white and pure red; there would be no pink. What is more, if we reversed the stirring, we would return to the perfect white/red separation. However, that cannot be—in a very few turns of the stirring stick the thickness of the layers goes from centimeters to Ångströms, and the result is irreversibly pink.

A century ago it seemed reasonable to assume that statistical mechanics applies only to systems with very many degrees of freedom. More recent is the

realization that much of statistical mechanics follows from chaotic dynamics, and already at the level of a few degrees of freedom the evolution of densities is irreversible. Furthermore, the theory that we shall develop here generalizes notions of ‘measure’ and ‘averaging’ to systems far from equilibrium, and transports us into regions hitherto inaccessible with the tools of equilibrium statistical mechanics.

By going to a description in terms of the asymptotic time evolution operators we give up tracking individual trajectories for long times, but trade in the uncontrollable trajectories for a powerful description of the asymptotic trajectory densities. This will enable us, for example, to give exact formulas for transport coefficients such as the diffusion constants without *any* probabilistic assumptions. The classical Boltzmann equation for evolution of 1-particle density is based on *stosszahlansatz*, neglect of particle correlations prior to, or after a 2-particle collision. It is a very good approximate description of dilute gas dynamics, but a difficult starting point for inclusion of systematic corrections. In the theory developed here, no correlations are neglected - they are all included in the cycle averaging formulas such as the cycle expansion for the diffusion constant $2dD = \lim_{T \rightarrow \infty} \langle x(T)^2 \rangle / T$ of a particle diffusing chaotically across a spatially-periodic array,

chapter 24

section 24.1

$$D = \frac{1}{2d} \frac{1}{\langle T \rangle_\zeta} \sum' (-1)^{k+1} \frac{(\hat{n}_{p_1} + \cdots + \hat{n}_{p_k})^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}, \quad (1.18)$$

where $\hat{n}_p$ is a translation along one period of a spatially periodic ‘runaway’ trajectory p . Such formulas are *exact*; the issue in their applications is what are the most effective schemes of estimating the infinite cycle sums required for their evaluation. Unlike most statistical mechanics, here there are no phenomenological macroscopic parameters; quantities such as transport coefficients are calculable to any desired accuracy from the microscopic dynamics.

The concepts of equilibrium statistical mechanics do help us, however, to understand the ways in which the simple-minded periodic orbit theory falters. A nonhyperbolicity of the dynamics manifests itself in power-law correlations and even ‘phase transitions.’

chapter 29

1.8 Chaos: what is it good for?

Happy families are all alike; every unhappy family is unhappy in its own way.

— *Anna Karenina*, by Leo Tolstoy

With initial data accuracy $\delta x = |\delta \mathbf{x}(0)|$ and system size L , a trajectory is predictable only up to the *finite* Lyapunov time (1.1), $T_{\text{Lyap}} \approx \lambda^{-1} \ln |L/\delta x|$. Beyond that, chaos rules. And so the most successful applications of ‘chaos theory’ have so far been to problems where observation time is much longer than a typical ‘turnover’ time, such as statistical mechanics, quantum mechanics, and questions of long term stability in celestial mechanics, where the notion of tracking accurately a given state of the system is nonsensical.

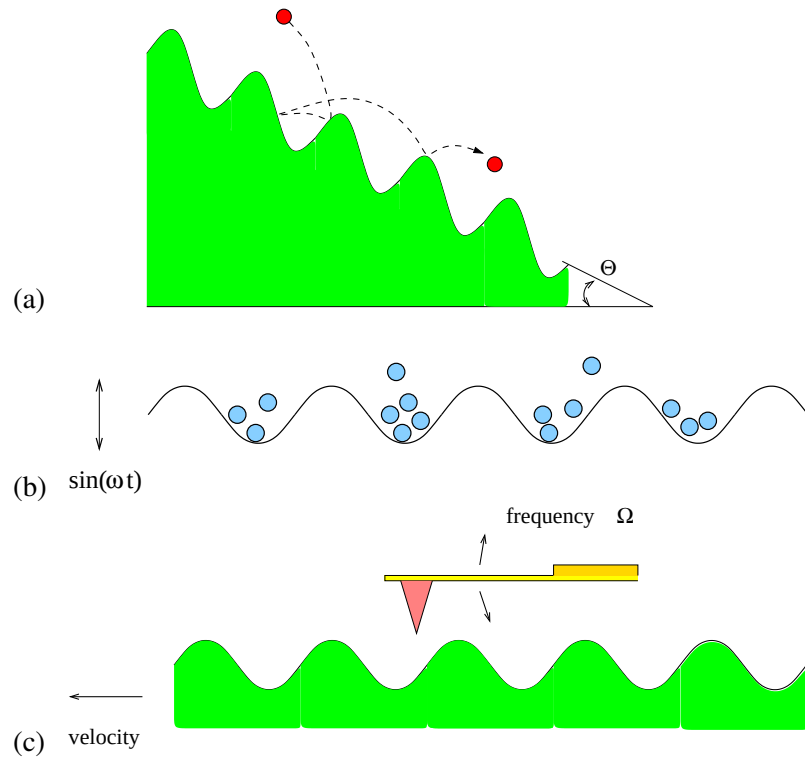


Figure 1.13: (a) Washboard mean velocity, (b) cold atom lattice diffusion, and (c) AFM tip drag force. (Y. Lan)

So what is chaos good for? *Transport!* Though superficially indistinguishable from the probabilistic random walk diffusion, in low dimensional settings the deterministic diffusion is quite recognizable, through the fractal dependence of the diffusion constant on the system parameters, and perhaps through non-Gaussian relaxation to equilibrium (non-vanishing Burnett coefficients).

section 24.2.1

Several tabletop experiments that could measure transport on macroscopic scales are sketched in figure 1.13 (each a tabletop, but an expensive tabletop). Figure 1.13 (a) depicts a ‘tilted washboard,’ a particle in a gravity field bouncing down the washboard, losing some energy at each bounce, or a charged particle in a constant electric field trickling across a periodic condensed-matter device. The interplay between chaotic dynamics and energy loss results in a terminal mean velocity/conductance, a function of the washboard tilt or external electric field that the periodic theory can predict accurately. Figure 1.13 (b) depicts a ‘cold atom lattice’ of very accurate spatial periodicity, with a dilute cloud of atoms placed onto a standing wave established by strong laser fields. Interaction of gravity with gentle time-periodic jiggling of the EM fields induces a diffusion of the atomic cloud, with a diffusion constant predicted by the periodic orbit theory. Figure 1.13 (c) depicts a tip of an atomic force microscope (AFM) bouncing against a periodic atomic surface moving at a constant velocity. The frictional drag experienced is the interplay of the chaotic bouncing of the tip and the energy loss at each tip/surface collision, accurately predicted by the periodic orbit theory. None of these experiments have actually been carried out, (save for some numerical experimentation), but are within reach of what can be measured today.

remark 24.3

Given microscopic dynamics, periodic orbit theory predicts observable macro-

scopic transport quantities such as the washboard mean velocity, cold atom lattice diffusion constant, and AFM tip drag force. But the experimental proposal is sexier than that, and goes into the heart of dynamical systems theory.

remark A1.1

Smale 1960s theory of the hyperbolic structure of the non-wandering set (AKA ‘horseshoe’) was motivated by his ‘structural stability’ conjecture, which - in non-technical terms - asserts that all trajectories of a chaotic dynamical system deform smoothly under small variations of system parameters.

Why this cannot be true for a system like the washboard in figure 1.13 (a) is easy to see for a cyclist. Take a trajectory which barely grazes the tip of one of the grooves. An arbitrarily small change in the washboard slope can result in loss of this collision, change a forward scattering into a backward scattering, and lead to a discontinuous contribution to the mean velocity. You might hold out hope that such events are rare and average out, but not so - a loss of a short cycle leads to a significant change in the cycle-expansion formula for a transport coefficient, such as (1.18).

When we write an equation, it is typically parameterized by a set of parameters by as coupling strengths, and we think of dynamical systems obtained by a smooth variation of a parameter as a ‘family.’ We would expect measurable predictions to also vary smoothly, i.e., be ‘structurally stable.’

But dynamical systems families are ‘families’ only in a name. That the structural stability conjecture turned out to be badly wrong is, however, not a blow for chaotic dynamics. Quite to the contrary, it is actually a virtue, perhaps the most dramatic experimentally measurable prediction of chaotic dynamics.

section 15.2

As long as microscopic periodicity is exact, the prediction is counterintuitive for a physicist - transport coefficients are *not* smooth functions of system parameters, rather they are non-monotonic, *nowhere differentiable* functions. Conversely, if the macroscopic measurement yields a smooth dependence of the transport on system parameters, the periodicity of the microscopic lattice is degraded by impurities, and probabilistic assumptions of traditional statistical mechanics apply. So the proposal is to –by measuring *macroscopic transport*– conductance, diffusion, drag –observe determinism on *nanoscales*, and –for example– determine whether an atomic surface is clean.

section 24.2

chapter 30

The signatures of deterministic chaos are even more baffling to an engineer: a small increase of pressure across a pipe exhibiting turbulent flow does not necessarily lead to an increase in the mean flow; mean flow dependence on pressure drop across the pipe is also a fractal function.

Is this in contradiction with the traditional statistical mechanics? No - deterministic chaos predictions are valid in settings where a few degrees of freedom are important, and chaotic motion time and space scales are commensurate with the external driving and spatial scales. Further degrees of freedom act as noise that smooths out the above fractal effects and restores a smooth functional dependence of transport coefficients on external parameters.

1.9 What is not in ChaosBook

There is only one thing which interests me vitally now, and that is the recording of all that which is omitted in books. Nobody, as far as I can see, is making use of those elements in the air which give direction and motivation to our lives.

— Henry Miller, *Tropic of Cancer*

This book offers everyman a breach into a domain hitherto reputed unreachable, a domain traditionally traversed only by mathematical physicists and mathematicians. What distinguishes it from mathematics is the insistence on computability and numerical convergence of methods offered. A rigorous proof, the end of the story as far as a mathematician is concerned, might state that in a given setting, for times in excess of 10^{32} years, turbulent dynamics settles onto an attractor of dimension less than 600. Such a theorem is of a little use to a hard-working plumber, especially if her hands-on experience is that within the span of a few typical ‘turnaround’ times the dynamics seems to settle on a (transient?) attractor of dimension less than 3. If rigor, magic, fractals or brains is your thing, read remark 1.5 and beyond.

So, no proofs! but lot of hands-on plumbing ahead.

Many a chapter alone could easily grow to a book size if unchecked: the nuts and bolt of the theory include ODEs, PDEs, stochastic ODEs, path integrals, group theory, coding theory, graph theory, ergodic theory, linear operator theory, quantum mechanics, etc.. We include material into the text proper on ‘need-to-know’ basis, relegate technical details to appendices, and give pointers to further reading in the remarks at the end of each chapter.

Résumé

This text is an exposition of the best of all possible theories of deterministic chaos, and the strategy is: 1) count, 2) weigh, 3) add up.

In a chaotic system any open ball of initial conditions, no matter how small, will spread over the entire accessible state space. Hence the theory focuses on describing the geometry of the space of possible outcomes, and evaluating averages over this space, rather than attempting the impossible: precise prediction of individual trajectories. The dynamics of densities of trajectories is described in terms of evolution operators. In the evolution operator formalism the dynamical averages are given by exact formulas, extracted from the spectra of evolution operators. The key tools are *trace formulas* and *spectral determinants*.

The theory of evaluation of the spectra of evolution operators presented here is based on the observation that the motion in dynamical systems of few degrees of freedom is often organized around a few *fundamental* cycles. These short cycles

capture the skeletal topology of the motion on a strange attractor/repeller in the sense that any long orbit can approximately be pieced together from the nearby periodic orbits of finite period. This notion is made precise by approximating orbits by prime cycles, and evaluating the associated curvatures. A curvature measures the deviation of a longer cycle from its approximation by shorter cycles; smoothness and the local instability of the flow implies exponential (or faster) fall-off for (almost) all curvatures. Cycle expansions offer an efficient method for evaluating deterministic and quantum observables.

The critical step in the derivation of the dynamical zeta function was the hyperbolicity assumption, i.e., the assumption of exponential shrinkage of all strips of the pinball repeller. By dropping the a_i prefactors in (1.5), we have given up on any possibility of recovering the precise distribution of starting x (which should anyhow be impossible due to the exponential growth of errors), but in exchange we gain an effective description of the asymptotic behavior of the system. The pleasant surprise of cycle expansions (1.10) is that the infinite time behavior of an unstable system is as easy to determine as the short time behavior.

To keep the exposition simple we have here illustrated the utility of cycles and their curvatures by a pinball game, but topics covered in ChaosBook – unstable flows, Poincaré sections, Smale horseshoes, symbolic dynamics, pruning, discrete symmetries, periodic orbits, averaging over chaotic sets, evolution operators, dynamical zeta functions, spectral determinants, cycle expansions, quantum trace formulas, zeta functions, and so on to the semiclassical quantization of helium – should give the reader some confidence in the broad sway of the theory. The formalism should work for any average over any chaotic set which satisfies two conditions:

1. the weight associated with the observable under consideration is multiplicative along the trajectory,
2. the set is organized in such a way that the nearby points in the symbolic dynamics have nearby weights.

The theory is applicable to evaluation of a broad class of quantities characterizing chaotic systems, such as the escape rates, Lyapunov exponents, transport coefficients and quantum eigenvalues. A big surprise is that the semiclassical quantum mechanics of systems deterministically chaotic is very much like the deterministic mechanics of chaotic systems; both are described by zeta functions and cycle expansions of the same form, with the same dependence on the topology of the deterministic flow.

But the power of instruction is seldom of much efficacy,
except in those happy dispositions where it is almost superfluous.

—Gibbon

Commentary

Remark 1.1. Preparatory nonlinear dynamics texts. ChaosBook aims to bridge the gap between the physics and mathematics dynamical systems literature. The intended audience is Alice, the perfect physics graduate student with a theoretical bent who does not believe anything she is told. As a complementary presentation we recommend Gaspard’s monograph [14] which covers much of the same ground in a highly readable and scholarly manner.

As far as the prerequisites are concerned—ChaosBook is not an introduction to nonlinear dynamics. Nonlinear science requires a one semester basic course (advanced undergraduate or first year graduate). A good start is the textbook by Strogatz [45], an introduction to the applied mathematician’s visualization of flows, fixed points, manifolds, bifurcations. It is the most accessible introduction to nonlinear dynamics—a book on differential equations in nonlinear disguise, and its broadly chosen examples and many exercises make it a favorite with students. It is not strong on chaos. There the textbook of Alligood, Sauer and Yorke [2] is preferable: an elegant introduction to maps, chaos, period doubling, symbolic dynamics, fractals, dimensions—a good companion to ChaosBook. Introduction more comfortable to physicists is the textbook by Ott [30], with the baker’s map used to illustrate many key techniques in analysis of chaotic systems. Ott is perhaps harder than the above two as the first textbook on nonlinear dynamics. Sprott [44] and Jackson [19] textbooks are very useful compendia of the ’70s and onward ‘chaos’ literature which we, in the spirit of promises made in sect. 1.1, tend to pass over in silence.

An introductory course should give students skills in qualitative and numerical analysis of dynamical systems for short times (trajectories, fixed points, bifurcations) and familiarize them with Cantor sets and symbolic dynamics for chaotic systems. For the dynamical systems material covered here in chapters 2 to 4, as well as for the in-depth study of bifurcation theory we warmly recommend Kuznetsov [24]. A good introduction to numerical experimentation with physically realistic systems is Tufillaro, Abbott, and Reilly [47]. Willis [49] short course on *Equilibria, periodic orbits and computing them*, is a very nice, student friendly introduction to numerical methods that underpin modern fluid-dynamical applications of the theory developed in ChaosBook, with online code samples. Korsch and Jodl [23] and Nusse and Yorke [29] also emphasize hands-on approach to dynamics.

There are many excellent online introductory chaos courses. For some, see ChaosBook.org/course1/preparatory.html. Robert Ghrist’s mesmerizing [silent movies](#) are perfect complement to them.

With any of the above, and a graduate level-exposure to statistical mechanics, partial differential equations and quantum mechanics, the stage is set for any of the one-semester advanced courses based on ChaosBook.

Question 1.2. Alice asks

Q You do not do *bifurcations*?

A No, we do not do bifurcations here. You should already know all about bifurcations.

Remark 1.2. ChaosBook based courses. The best ChaosBook course is the ChaosBook.org/course1 course. Other courses taught so far (for a listing, consult ChaosBook.org/courses) start out with the introductory chapters on qualitative dynamics, symbolic dynamics and flows, and then continue in different directions:

Deterministic chaos. Chaotic averaging, evolution operators, trace formulas, zeta functions, cycle expansions, Lyapunov exponents, billiards, transport coefficients, thermody-

dynamic formalism, period doubling, renormalization operators. For a quick tour, take [Omri Gat's 2015 *Periodic orbit theory of chaos*](#) (all of it in a [Mathematica notebook](#)) for a spin. A graduate level introduction to statistical mechanics from the dynamical point view is given by Dorfman [11]; the Gaspard monograph [14] covers the same ground in more depth. Driebe monograph [12] offers a nice introduction to the problem of irreversibility in dynamics. The role of ‘chaos’ in statistical mechanics is critically dissected by Bricmont in his highly readable essay “*Science of Chaos or Chaos in Science?*” [10].

Spatiotemporal dynamical systems. Partial differential equations for dissipative systems, weak amplitude expansions, normal forms, symmetries and bifurcations, pseudospectral methods, spatiotemporal chaos, turbulence. Holmes, Lumley and Berkooz [18] offer a delightful discussion of why the Kuramoto-Sivashinsky equation deserves study as a staging ground for a dynamical approach to study of turbulence in full-fledged Navier-Stokes boundary shear flows (consult chapter 30).

Quantum chaos. Semiclassical propagators, density of states, trace formulas, semiclassical spectral determinants, billiards, semiclassical helium, diffraction, creeping, tunneling, higher-order $\hbar$ corrections. To read about this, hop to the quantum chaos introduction, chapter 35.

chapter 35

Question 1.3. Alice asks

Q Why do they always teach classical mechanics first, nonlinear dynamics second? Isn't that upside down?

A Beats me.

Remark 1.3. ‘Strange attractor’. A ‘strange attractor’ appeared for the first time [38, 39] in a 1971 David Ruelle and Floris Takens article [37]. The article was meant to clarify a small mathematical point concerning hydrodynamic turbulence, and authors assumed the article would go unnoticed and be immediately forgotten. Instead, it went viral, becoming one of most cited publications in nonlinear science.

Remark 1.4. Periodic orbit theory. This book puts more emphasis on periodic orbit theory than any other current nonlinear dynamics textbook. The role of unstable periodic orbits was already fully appreciated by Poincaré [8, 33], who noted that hidden in the apparent chaos is a rigid skeleton, a tree of *cycles* (periodic orbits) of increasing periods and self-similar structure, and suggested that the cycles should be the key to chaotic dynamics. Periodic orbits have been at core of much of the mathematical work on the theory of the deterministic and quantum dynamical systems ever since. We refer the reader to the reprint selection [26] for an overview of some of that literature.

Remark 1.5. If you seek rigor? If you find ChaosBook not rigorous enough, you should turn to the mathematics literature. In order to aid the reader in digesting this literature, we try to point out the notational correspondences whenever appropriate. The formalism is replete with lims, sups, infs, Ω -sets, etc. Whenever we deem such rigor not essential to understanding the theory, we avoid it. We give a short shrift to the theory of bifurcations, and the KAM (Kolmogorov-Arnol'd-Moser) tori make only a tangential appearance. We recommend Robinson's advanced graduate level exposition of dynamical systems theory [34] from Smale perspective. The most extensive reference is the treatise by Katok and Hasselblatt [21], an impressive compendium of modern dynamical systems theory. The fundamental papers in this field, all still valuable reading, are Smale [43], Bowen [9] and Sinai [42]. Sinai's paper is prescient and offers a vision and a program that ties together dynamical systems and statistical mechanics. It is written

for readers versed in statistical mechanics. For a dynamical systems exposition, consult Anosov and Sinai [3]. Markov partitions were introduced by Sinai in ref. [41]. The classical text (though certainly not an easy read) on the subject of dynamical zeta functions is Ruelle's *Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics* [36]. In Ruelle's monograph transfer operator technique (or the 'Perron-Frobenius theory') and Smale's theory of hyperbolic flows are applied to zeta functions and correlation functions. The status of the theory from Ruelle's point of view is compactly summarized in his 1995 Pisa lectures [35]. Further excellent mathematical references on thermodynamic formalism are Parry and Pollicott's monograph [31] with emphasis on the symbolic dynamics aspects of the formalism, and Baladi's clear and compact reviews of the theory of dynamical zeta functions [5, 6].

Remark 1.6. If you seek magic? ChaosBook resolutely skirts number-theoretical magic such as spaces of constant negative curvature, Poincaré tilings, modular domains, Selberg Zeta functions, Riemann hypothesis, ... Why? While this beautiful mathematics has been very inspirational, especially in studies of quantum chaos, almost no powerful method in its repertoire survives a transplant to a physical system that you are likely to care about.

Remark 1.7. Grasshoppers vs. butterflies. The 'sensitivity to initial conditions' was discussed by Maxwell, then 30 years later by Poincaré. In weather prediction, the Lorenz 'Butterfly Effect' started its journey in 1898, as a 'Grasshopper Effect' in a book review by W. S. Franklin [13]. In 1963 Lorenz ascribed a 'seagull effect' to an unnamed meteorologist, and in 1972 he repackaged it as the 'Butterfly Effect'. Jamie L. Vernon [48] writes: "During the 139th meeting of the American Association for the Advancement of Science, Edward Lorenz posed a question, 'Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?' Lorenz's insight called into question laws introduced as early as 1687 by Sir Isaac Newton suggesting that nature is a probabilistic mechanical system, 'a clockwork universe.' Similarly, Lorenz challenged Pierre-Simon Laplace, who argued that unpredictability has no place in the universe, asserting that if we knew all the physical laws of nature, then 'nothing would be uncertain and the future, as the past, would be present to [our] eyes.' " In this video  Sabine Hossenfelder explains what Lorenz really meant by the 'Butterfly Effect'.

remark 2.3

Remark 1.8. Sorry, no schmactals!

After all, it's impossible to read a single tweet, or hear him speak a sentence or two, without staring deep into the abyss. He turns being artless into an art form; he is a Picasso of pettiness; a Shakespeare of s**t. His faults are fractal: even his flaws have flaws, and so on ad infinitum.

—Nate White

Question 1.4. Alice asks

Q Before any serious study of the topic, fractals would have been the first word to come to my mind at the mention of chaos theory. So, if I may, why are fractals on the outs?

A We try to explain why in sect. 1.3.3: it's a regrettable historical accident – fractal pictures are cute, but not how the theory of chaotic dynamics actually works, which is a subject much deeper and intellectually more beautiful – hence ChaosBook. Basically, in the 1980's physicist were trying to learn the new subject, and spent much time on 1-, 2-, 3-dimensional systems that they could visualize playing with computers. Some insights were fruitful in understanding high-dimensional, physical problems. Fractals were not one of them.

section 1.3.3

On a hype-free planet, the totality of what Hale & Koçak [17] have to say about this baby-boomer phenomenon would suffice: “No exposition of planar maps would be complete without mentioning fractals; so we mention them. Some of the popular resources are Barnsley [7] and Peitgen & Richter [32].”

But, as people ask, we must say something about them. The word ‘fractal’ was coined by Mandelbrot [27]. Addison’s introduction to fractal dimensions [1] offers a well-motivated entry into this field. ChaosBook skirts mathematics and empirical practice of fractal analysis, such as Hausdorff and fractal dimensions. For reasons that remain mysterious to the authors - perhaps so that Mandelbrot could refer to himself both as the mother of fractals and the grandmother of multifractals - some physics literature refers to any fractal generated by more than one scale as a ‘multi’-fractal. This usage divides fractals into 2 classes; one consisting of the canonical $1/3$ ’s Cantor set and the Serapinski gasket, and the second consisting of anything else, including all cases of physical interest. A bit like naming all one-legged creatures ‘monopeds’, and then claiming the credit for the sole discovery of all two- or more long-legged beasts, and claiming the honor of naming them ‘multipeds’. Even though the experimental evidence for the fractal geometry of nature is circumstantial [4], in studies of probabilistically assembled fractal aggregates such as diffusion limited aggregates (DLA) better measures of ‘complexity’ are lacking. For deterministic systems, however, we can do *much* better, by studying physically motivated and experimentally measurable quantities (escape rates, diffusion coefficients, energy dissipation rates of turbulent flows, semiclassical atomic spectra, ...). That’s what the ChaosBook is about.

remark A1.5

Remark 1.9. Dynamics *is*! This comes up a lot, so might just as well dispose of it right away. “Dynamics is,” not “Dynamics are:”

dy·nam·ics (used with a singular verb) The branch of mechanics that deals with the motion and equilibrium of systems under the action of forces, usually from outside the system.

Economist style guide says:

“A government, a party, [...] are all it and take a singular verb. So does a country, even if its name looks plural. Thus The Philippines has a congressional system, as does the United States; the Netherlands does not. The United Nations is also singular. So are acoustics, ballistics, dynamics, economics, kinetics, mathematics, mechanics, physics, politics and statics when being used generally, without the definite article. But such -ics words are plural when preceded by the, or the plus an adjective, or with a possessive. [...] ‘The dynamics of the dynasty were dynamite’...

Remark 1.10. Rat brains? If you were wondering while reading this introduction ‘what’s up with rat brains?’, the answer is yes indeed, there is a line of research in neuronal dynamics that focuses on possible unstable periodic states, described for example in refs. [15, 16, 28, 40].

A guide to exercises

God can afford to make mistakes. So can Dada!
—Dadaist Manifesto

The essence of this subject is incommunicable in print; the only way to develop intuition about chaotic dynamics is by computing, and the reader is urged to try to work through the essential exercises. As not to fragment the text, the exercises are indicated by text margin boxes such as the one on this margin, and collected at the end of each chapter. By the end of a (two-semester) course you should have completed at least three small projects: (a) compute everything for a 1-dimensional repeller, (b) compute escape rate for a 3-disk game of pinball, (c) compute a part of the quantum 3-disk game of pinball, or the helium spectrum, or if you are interested in statistical rather than the quantum mechanics, compute a transport coefficient. The essential steps are:

exercise 23.2

- **Dynamics**

1. count prime cycles, exercise 1.1, exercise 11.3, exercise 14.1
2. pinball simulator, exercise 9.1, exercise 16.4
3. pinball stability, exercise 16.6, exercise 16.4
4. pinball periodic orbits, exercise 16.5, exercise 16.3
5. helium integrator, exercise 2.11, exercise 7.4
6. helium periodic orbits, exercise 16.10

- **Averaging, numerical**

1. pinball escape rate, exercise 20.3

- **Averaging, periodic orbits**

1. cycle expansions, exercise 23.1, exercise 23.2
2. pinball escape rate, exercise 23.4, exercise 23.5
3. cycle expansions for averages, exercise 23.1, exercise 27.3
4. cycle expansions for diffusion, exercise 24.1
5. pruning, transition graphs, exercise 18.6
6. desymmetrization exercise 25.1
7. intermittency, phase transitions, exercise 29.6

The exercises that you should do have **underlined titles**. The rest (**smaller type**) are optional. Difficult problems are marked by a chili 🌶️, or any number of *** stars. If you solve one of those, it is probably worth a [publication](#). Solutions to many of the problems are available upon request. A clean solution, a pretty figure, or a nice exercise that you contribute to ChaosBook will be gratefully acknowledged. Often going through a solution is more instructive than reading the chapter that problem is supposed to illustrate.

References

- [1] P. S. Addison, *Fractals and Chaos: An illustrated Course* (CRC Press, Boca Raton FL, 1997).
- [2] K. T. Alligood, T. D. Sauer, and J. A. Yorke, *Chaos, An Introduction to Dynamical Systems* (Springer, New York, 1996).
- [3] D. V. Anosov and Y. G. Sinai, “Some smooth ergodic systems”, *Russ. Math. Surv.* **22**, 103–167 (1967).
- [4] D. Avnir, O. Biham, D. Lidar, and O. Malcai, “Is the geometry of nature fractal?”, *Science* **279**, 39–40 (1998).
- [5] V. Baladi, Dynamical zeta functions, in *Real and Complex Dynamical Systems: Proceedings of the NATO ASI*, edited by B. Branner and P. Hjorth (1995), pp. 1–26.
- [6] V. Baladi, *Positive Transfer Operators and Decay of Correlations* (World Scientific, Singapore, 2000).
- [7] M. F. Barnsley, *Fractals Everywhere* (Dover, New York, 2012).
- [8] J. Barrow-Green, *Poincaré and the Three Body Problem* (Amer. Math. Soc., Providence R.I., 1997).
- [9] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms* (Springer, New York, 1975).
- [10] J. Bricmont, “Science of chaos or chaos in science?”, *Ann. New York Acad. Sci.* **775**, 131–175 (1995).
- [11] J. R. Dorfman, *An Introduction to Chaos in Nonequilibrium Statistical Mechanics* (Cambridge Univ. Press, Cambridge UK, 1999).
- [12] D. J. Driebe, *Fully Chaotic Maps and Broken Time Symmetry* (Springer, New York, 1999).
- [13] W. S. Franklin, “New books”, *Phys. Rev.* **6**, 173–174 (1898).
- [14] P. Gaspard, *Chaos, Scattering and Statistical Mechanics* (Cambridge Univ. Press, Cambridge UK, 1998).
- [15] J. Glanz, “Do chaos-control techniques offer hope for epilepsy?”, *Science* **265**, 1174 (1994).
- [16] J. Glanz, “Mastering the nonlinear brain”, *Science* **277**, 1758–1760 (1997).
- [17] J. Hale and H. Koçak, *Dynamics and Bifurcations* (Springer, New York, 1991).
- [18] P. Holmes, J. L. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge Univ. Press, Cambridge UK, 1996).
- [19] E. A. Jackson, *Perspectives of Nonlinear Dynamics*, Vol. 1 (Cambridge Univ. Press, Cambridge UK, 1989).
- [20] K. Jones-Smith and H. Mathur, “Fractal analysis: Revisiting Pollock’s drip paintings”, *Nature* **444**, E9–E10 (2006).

- [21] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems* (Cambridge Univ. Press, Cambridge UK, 1995).
- [22] R. Kennedy, *The case of Pollock's fractals focuses on physics*, New York Times, Dec. 2, 2006.
- [23] H. J. Korsch and H.-J. Jodl, *Chaos: A Program Collection for the PC* (Springer, New York, 1994).
- [24] Y. A. Kuznetsov, *Elements of Applied Bifurcation Theory*, 3rd ed. (Springer, New York, 2004).
- [25] G. W. Leibniz, *Hauptschriften zur Grundlegung der Philosophie*, in (Ernst Cassirer, Leibzig, 1906) Chap. Von dem Verhängnisse, pp. 129–134.
- [26] R. S. MacKay and J. D. Meiss, *Hamiltonian Dynamical Systems: A Reprint Selection* (Taylor & Francis, Oxford, 1987).
- [27] B. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).
- [28] F. Moss, “Chaos under control”, *Nature* **370**, 596–597 (1994).
- [29] H. E. Nusse and J. A. Yorke, *Dynamics: Numerical Explorations* (Springer, New York, 1997).
- [30] E. Ott, *Chaos and Dynamical Systems* (Cambridge Univ. Press, Cambridge UK, 2002).
- [31] W. Parry and M. Pollicott, *Zeta Functions and the Periodic Orbit Structure of Hyperbolic Dynamics*, Vol. 187–188 (Astérisque, Soc. Math. France, 1990).
- [32] H. O. Peitgen and P. H. Richter, *The Beauty of Fractals. Images of Complex Dynamical Systems* (Springer, Berlin, 1986).
- [33] H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste* (Guthier-Villars, Paris, 1899).
- [34] C. Robinson, *Dynamical Systems: Stability, Symbolic Dynamics, and Chaos* (CRC Press, Boca Raton FL, 1995).
- [35] D. Ruelle, *Functional determinants related to dynamical systems and the thermodynamic formalism*, tech. rep. IHES-P-95-30 (Inst. Hautes Etud. Sci., Bures-sur-Yvette, 1995).
- [36] D. Ruelle, *Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics*, 2nd ed. (Cambridge Univ. Press, Cambridge UK, 2004).
- [37] D. Ruelle and F. Takens, “On the nature of turbulence”, *Commun. Math. Phys.* **20**, 167 (1971).
- [38] H. H. Rugh, “Un interview de David Ruelle”, *Gaz. Math.* **161**, 55–59 (2019).
- [39] H. H. Rugh, “An interview with David Ruelle”, *EMS Newsletter* **2020-3**, 21–24 (2020).
- [40] S. J. Schiff, K. Jerger, D. H. Duong, T. Chang, M. L. Spano, and W. L. Ditto, “Controlling chaos in the brain”, *Nature* **370**, 615–620 (1994).

- [41] Y. G. Sinai, “Construction of Markov partitions”, *Funct. Anal. Appl.* **2**, 245–253 (1968).
- [42] Y. G. Sinai, “Gibbs measures in ergodic theory”, *Russ. Math. Surv.* **27**, 21–69 (1972).
- [43] S. Smale, “Differentiable dynamical systems”, *Bull. Amer. Math. Soc.* **73**, 747–817 (1967).
- [44] J. C. Sprott, *Chaos and Time-series Analysis* (Oxford Univ. Press, Oxford, 2003).
- [45] S. H. Strogatz, *Nonlinear Dynamics and Chaos* (Westview Press, Boulder, CO, 2014).
- [46] R. P. Taylor, A. P. Micolich, and D. Jonas, “Fractal analysis of Pollock’s drip paintings”, *Nature* **399**, 422–422 (1999).
- [47] N. Tufillaro, T. Abbott, and J. Reilly, *An Experimental Approach to Non-linear Dynamics and Chaos* (Addison-Wesley, Redwood City, 1992).
- [48] J. L. Vernon, “Understanding the butterfly effect”, *Amer. Sci.* **105**, 130 (2017).
- [49] A. P. Willis, *Equilibria, periodic orbits and computing them*, 2019.

Exercises

- 1.1. **3-disk symbolic dynamics.** As periodic trajectories will turn out to be our main tool to breach deep into the realm of chaos, it pays to start familiarizing oneself with them now by sketching and counting the few shortest prime cycles (we return to this in sect. 18.4). Show that the 3-disk pinball has $3 \cdot 2^{n-1}$ itineraries of length n . List periodic orbits of periods 2, 3, 4, 5, $\dots$. Verify that the shortest 3-disk prime cycles are 12, 13, 23, 123, 132, 1213, 1232, 1323, 12123, $\dots$. Try to sketch them. (continued in exercise 15.7)
- 1.2. **Sensitivity to initial conditions.** Assume that two pinball trajectories start out parallel, but separated by 1 Ångström, and the disks are of radius $a = 1$ cm and center-to-center separation $R = 6$ cm. Try to estimate in how many bounces the separation will grow to the size of system (assuming that the trajectories have been picked so they remain trapped for at least that long). Estimate the Who's *Pinball Wizard*'s typical score (number of bounces) in a game without cheating, by hook or crook (by the end of chapter 23 you should be in position to make very accurate estimates).

Chapter 2

Go with the flow

Dynamical systems theory includes an extensive body of knowledge about qualitative properties of generic smooth families of vector fields and discrete maps. The theory characterizes structurally stable invariant sets [...] The logic of dynamical systems theory is subtle. The theory abandons the goal of describing the qualitative dynamics of all systems as hopeless and instead restricts its attention to phenomena that are found in selected systems. The subtlety comes in specifying the systems of interest and which dynamical phenomena are to be analyzed.

— John Guckenheimer

(R. Mainieri, P. Cvitanović and E.A. Spiegel)

WE DEFINE a *dynamical system* (M, f) and classify its solutions as equilibria, periodic, and aperiodic. An ‘aperiodic’ solution is either ‘wandering’ or belongs to a non-wandering set, which in turn can be decomposed into chain-recurrent sets. Various cases are illustrated with concrete examples, such as the Rössler and Lorenz systems.



fast track:
chapter 19, p. 381

2.1 Dynamical systems

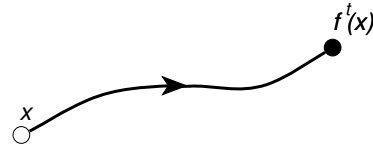
I would have written a shorter book, but I didn’t have the time.

— Channeling Blaise Pascal

In a dynamical system we observe the world as it evolves with time. We express



Figure 2.1: A trajectory traced out by the evolution rule f^t . Starting from the state space point x , after a time t , the point is at $f^t(x)$.



our observations as numbers and record how they change; given sufficiently detailed information and understanding of the underlying natural laws, we see the future in the present as in a mirror. The motion of the planets against the celestial firmament provides an example. As the fixed stars traverse the sky from East to West, the planets distinguish themselves by moving relative to the fixed stars. Ancients discovered that by knowing a sequence of planet's positions—latitudes and longitudes—its future position could be predicted. ¹

section 1.3

For the solar system, tracking the latitude and longitude in the celestial sphere suffices to completely specify the planet's apparent motion. All possible values for positions and velocities of the planets form the *phase space* of the system. More generally, a state of a physical system, at a given instant in time, can be represented by a single point in an abstract space called *state space* $\mathcal{M}$ (mnemonic: curly 'M' for a 'manifold'). As the system changes, so does the *representative point* in state space. We refer to the evolution of the totality of such points as a *flow* or *dynamics*. The function f^t , which specifies the representative point at time t , is the *evolution rule*.

remark 2.1

If there is a definite rule f that tells us how this representative point moves in $\mathcal{M}$, the system is said to be *deterministic*. For a deterministic dynamical system, the evolution rule takes one point of the state space and maps it into exactly one point. However, this is not always possible. For example, knowing the temperature today is not enough to predict the temperature tomorrow; knowing the value of a stock today will not determine its value tomorrow. With a sufficiently large state space, we might be able to formulate an evolution rule. For example, this expanded state space allows us to predict tomorrow's temperature by measuring points over the entire atmosphere of the planet. Even this is not quite true, and we are less hopeful when it comes to stocks.

For a deterministic system almost every point has a unique future, so trajectories cannot intersect. We say 'almost' because there might exist a set of measure zero (tips of wedges, cusps, etc.) for which a trajectory is not defined. We may think such sets a nuisance, but it is quite the contrary—they will enable us to partition state space, so that the dynamics can be better understood. ²



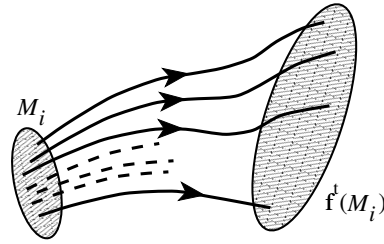
chapter 15

Dynamical *evolution* is an initial value problem, with d numbers sufficient to determine what will happen time t later. Locally, the state space $\mathcal{M}$ looks like $\mathbb{R}^d$. The local linear vector space (tangent space) at any given state space point $x \in \mathcal{M}$ can be thought of as a 'chart' (however, we shall use this term in a more

¹In order not to interrupt the flow of exposition, the examples are always relegated to a separate section, here sect. 2.6. But to *understand* the exposition, you have to work through the examples.

²Measure zero? Think of rational numbers in a unit interval. They are dense (there is infinity of them in any finite length interval), but you can throw darts at them until cows come home, and – with probability 1 – never hit one.

Figure 2.2: A flow: The evolution rule f^t can be used to map a region M_i of the state space into the region $f^t(M_i)$.



restricted sense, only after the continuous time and continuous symmetries have been ‘quotiented out’, see sects. 3.1 and 13.1). Globally, the state space may be a more complicated *manifold* such as a torus, a cylinder, or some other smooth geometric object. By *manifold* we mean a smooth differentiable d -dimensional space which looks like $\mathbb{R}^d$ only locally. For example, the state space flow of an autonomous Hamiltonian system is confined to a curved constant energy hypersurface. When we need to stress that the dimension d of $\mathcal{M}$ is greater than one, we may refer to the point $x \in \mathcal{M}$ as x_i where $i = 1, 2, 3, \dots, d$. If the dynamics is described by a set of PDEs (partial differential equations), the state space is the infinite-dimensional function space, with a given instantaneous state or *field* $u = u(x)$ labeled by a set of continuous indices x . The evolution rule $f^t : \mathcal{M} \rightarrow \mathcal{M}$ tells us where the initial state x lands in $\mathcal{M}$ after the time interval t .

chapter 30

section 2.4

The pair $(\mathcal{M}, f)$ constitute a *dynamical system*.

The dynamical systems we will be studying are smooth. This is expressed mathematically by saying that the evolution rule f^t can be differentiated as many times as needed. Its action on a point x is sometimes indicated by $f(x, t)$ to remind us that f is really a function of two variables: the time and a point in state space. Note that time is relative rather than absolute, so only the time interval is necessary. This follows from the fact that a point in state space completely determines all future evolution, and to locate where it lands in the future it is not necessary to know anything besides the elapsed time interval. The time parameter can be a real variable ($t \in \mathbb{R}$), in which case the evolution is called a *flow*, or an integer ($t \in \mathbb{Z}$), in which case the evolution advances in discrete steps in time, given by *iteration* of a *map*. The evolution parameter need not be the physical time; for example, a time-stationary solution of a partial differential equation is parameterized by spatial variables. In such situations one talks of a ‘spatial profile’ rather than a ‘flow’.



Nature provides us with innumerable dynamical systems. They manifest themselves through their orbits: given a state x_0 at initial time t_0 , the

$$\text{flow map} \quad f^t : \quad x_0 \rightarrow x(x_0, t) \quad (2.1)$$

yields the state $x(t)$ time t later. This evolution rule traces out a sequence of points $x(t) = f^t(x_0)$, the *orbit* through the point $x_0 = x(0)$. We usually omit the x_0 label from $x(x_0, t)$. By extension, we also talk of the evolution of a region M_i of the state

space. The language of continuum mechanics is quite helpful in visualizing such deformations, not only in 3-dimensional space, but also in state spaces of arbitrary dimension. Consider a motion f from the *undeformed* (reference or initial) region (a ‘body’) $\mathcal{M}_i$ to the *deformed* (current or final) region $\mathcal{M}_f = f^t(\mathcal{M}_i)$. We may write the motion as a map

$$f^t : \mathcal{M}_i \rightarrow \mathcal{M}_f, \quad (2.2)$$

such that every x_0 in $\mathcal{M}_i$ is mapped to an $x = f^t(x_0)$ in $\mathcal{M}_f$, as in figure 2.2, where x denotes the state in the deformed region, and x_0 represents the state in the initial, undeformed region.

exercise 2.1

The set of points that belong to the infinite-time trajectory (2.1) of a given point x_0 is called the *orbit* of x_0 ;

section 10.3

section 13.2

$$\text{time orbit of } x_0 : \mathcal{M}_{x_0} \subset \mathcal{M} \quad (2.3)$$

we shall talk about forward orbits, backward orbits, periodic orbits (see figure 2.3), etc.. For a flow, a time orbit is a 1-dimensional continuous curve; for a map, it is a sequence of points. In ChaosBook ‘trajectory’ refers to a set of points or a curve segment traced out by $x(t)$ over a finite time interval t . ‘Orbit’ refers to the totality of states that can be reached from x_0 , with state space $\mathcal{M}$ stratified into a union of such orbits (each $\mathcal{M}_{x_0}$ labeled by a single point belonging to the set, $x_0 = x(0)$ for example). Under time evolution a trajectory segment is mapped into another trajectory segment, but points within an orbit are only shifted; the orbit considered as a set is unchanged. Hence an orbit is a *dynamically invariant* notion.



The central idea of ChaosBook is to replace the complicated, ergodic, asymptotic $t \rightarrow \infty$ dynamics by a systematic hierarchy of compact *time-invariant* sets or compact orbits (equilibria, periodic orbits, invariant tori, $\dots$).

2.1.1 A classification of possible motions?

Ah, yes, Judgie, everything will go away someday. It’s the waiting that’s so exquisitely wearing.



— Duke Ellington, to Robert Traver

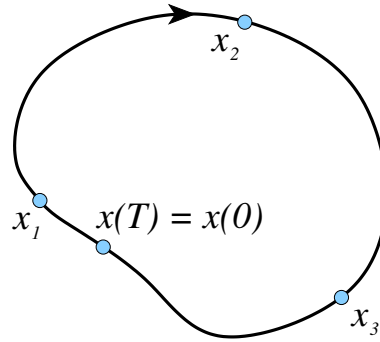
What kinds of orbits are there? This is a grand question, and there are many answers. Here is a first attempt to classify all possible orbits:



$$\begin{array}{lll} \text{stationary:} & f^t(x) = x & \text{for all } t \\ \text{periodic:} & f^t(x) = f^{t+T_p}(x) & \text{for a given minimum period } T_p \\ \text{aperiodic:} & f^t(x) \neq f^{t'}(x) & \text{for all } t \neq t' . \end{array}$$

A *periodic orbit* (or a *cycle*) p is the set of points $\mathcal{M}_p \subset \mathcal{M}$ swept out by a trajectory that returns to the initial point in a finite time. We refer to a point on a periodic orbit as a *periodic point*, see figure 2.3. Periodic orbits form a very small

Figure 2.3: A periodic point returns to the initial point after a finite time, $x = f^{T_p}(x)$. Periodic orbit p is the set of periodic points $p = \mathcal{M}_p = \{x_1, x_2, \dots\}$ swept out by the trajectory of any one of them in the finite time T_p .



subset of the state space, in the same sense that rational numbers are a set of zero measure on the unit interval.

chapter 5

Periodic orbits and equilibrium points are the simplest examples of ‘non-wandering’ invariant sets preserved by dynamics. Dynamics can also preserve higher-dimensional smooth compact invariant manifolds; most commonly encountered are the M -dimensional tori of Hamiltonian dynamics, with the notion of periodic motion generalized to quasiperiodic (the superposition of M incommensurate frequencies) motion on a smooth torus, and families of solutions related by a continuous symmetry. Further examples are afforded by stable / unstable manifolds (swept by semi-infinite curves originating at an equilibrium along each stability eigenvector) and the most baffling of all invariant orbits, the infinite time ergodic orbits.



section 15.1

The ancients tried to make sense of all dynamics in terms of periodic motions, epicycles, what we today call ‘integrable systems’. The embarrassing truth is that for a generic dynamical system almost all motions are aperiodic. So we refine the classification by dividing aperiodic motions into two subtypes: those that wander off, and those that keep coming back.

A point $x \in \mathcal{M}$ is called a *wandering point*, if there exists an open neighborhood $\mathcal{M}_0$ of x to which the orbit never returns

$$f^t(x) \notin \mathcal{M}_0 \quad \text{for all } t > t_{\min}. \quad (2.4)$$

In physics literature, the dynamics of such a state is often referred to as *transient*.



Wandering points do not take part in the long-time dynamics, so your first task is to prune them from $\mathcal{M}$ as well as you can. What remains envelops the set of the long-time orbits, or the *non-wandering set*.

For times much longer than a typical ‘turnover’ time, it makes sense to relax the notion of exact periodicity and replace it by the notion of *recurrence*. A point is *recurrent* or *non-wandering*, if for any open neighborhood $\mathcal{M}_0$ of x and any time $t_{\min}$ there exists a later time t , such that

$$f^t(x) \in \mathcal{M}_0. \quad (2.5)$$

In other words, the orbit of a non-wandering point reenters the neighborhood $\mathcal{M}_0$ infinitely often. We shall denote the *non-wandering set* of f by Ω , i.e., the

union of all the non-wandering points of $\mathcal{M}$. This non-wandering set of f is key to understanding the long-time behavior of a dynamical system; all calculations undertaken here will be carried out on non-wandering sets.



So much about individual trajectories. What about clouds of initial points? If there exists a connected state space volume that maps into itself under forward evolution (and you can prove that by the method of Lyapunov functionals, or several other methods available in the literature), the flow is globally contracting onto a subset of $\mathcal{M}$ which we shall refer to as the *attractor*. The attractor may be unique, or there can coexist any number of distinct attracting sets, each with its own *basin of attraction*, the set of all points that fall into the attractor under forward evolution. The attractor can be a fixed point (a *sink*), a periodic orbit (a *limit cycle*), aperiodic, or any combination of the above. The most interesting case is that of an aperiodic recurrent attractor, to which we shall refer loosely as a *strange attractor*. We say ‘loosely’, as will soon become apparent that diagnosing and proving existence of a genuine, card-carrying strange attractor is a highly nontrivial undertaking; it requires explaining notions like ‘transitive’ and ‘chain-recurrent’ that we will be ready to discuss only in sect. 17.1.



example 2.3

Conversely, a repeller is a non-wandering set Ω that is enclosed by a connected state space volume $\mathcal{M}_0$, where nearly all points eventually exit $\mathcal{M}_0$. An example of a repeller is not hard to come by—the pinball game of sect. 1.3 is a simple chaotic repeller. Ω , the *non-wandering set* of f , is the union of all invariant sets: attracting/repelling fixed points, strange attractors, repellers, etc..



It would seem, having said that the periodic points are so exceptional that almost all non-wandering points are aperiodic, that we have given up the ancients’ fixation on periodic motions. Nothing could be further from the truth. If longer and longer cycles approximate more and more accurately finite segments of aperiodic trajectories, we can establish control over non-wandering sets by defining them as the closure of the union of all periodic points.

Before we can work out an example of a non-wandering set and get a better grip on what chaotic motion might look like, we need to ponder flows in a little more depth.

2.2 Flows



Knowing the equations and knowing the solution are two different things. Far, far away.

— T.D. Lee

A *flow* is a continuous-time dynamical system. The evolution rule f^t is a family of mappings of $\mathcal{M} \rightarrow \mathcal{M}$ parameterized by $t \in \mathbb{R}$. Because t represents a time interval, any family of mappings that forms an evolution rule must satisfy:



exercise 2.2

- (a) $f^0(x) = x$ (in 0 time there is no motion)

- (b) $f^t(f^{t'}(x)) = f^{t+t'}(x)$ (the evolution law is the same at all times)
- (c) the mapping $(x, t) \mapsto f^t(x)$ from $\mathcal{M} \times \mathbb{R}$ into $\mathcal{M}$ is continuous.

We shall often find it convenient to represent functional composition by ‘ $\circ$ ’:

$$f^{t+s} = f^t \circ f^s = f^t(f^s). \quad (2.6)$$

appendix A10.1

The family of mappings $f^t(x)$ thus forms a continuous (1-parameter forward Lie semi-) group. Why ‘semi-’group? It may fail to form a group if the dynamics is not reversible, and the rule $f^t(x)$ cannot be used to rerun the dynamics backwards in time, with negative t ; with no reversibility, we cannot define the inverse $f^{-t}(f^t(x)) = f^0(x) = x$, in which case the family of mappings $f^t(x)$ does not form a group. The dynamics of an exceedingly large number of situations that lie beyond Lyapunov time cannot be run backwards, such as the dynamics of asymptotic attractors, dissipative partial differential equations, systems with noise, and non-invertible maps. Hence, there’s a circumspect emphasis on *semigroups*. On the other hand, there are many settings of physical interest, where dynamics is reversible (such as finite-dimensional Hamiltonian flows), and where the family of evolution maps f^t does form a group.

For infinitesimal times, flows can be defined by differential equations. We write a trajectory, a smooth curve embedded in the state space as

$$x(t + \tau) = f^{t+\tau}(x_0) = f(f(x_0, t), \tau) \quad (2.7)$$

and express the tangent to the curve at point $x(t)$ as

exercise 2.3

$$\left. \frac{dx}{d\tau} \right|_{\tau=0} = \partial_\tau f(f(x_0, t), \tau)|_{\tau=0} = \dot{x}(t), \quad (2.8)$$

the time derivative of the evolution rule, a vector evaluated at the point $x(t)$. By considering all possible orbits, we obtain the vector $\dot{x}(t)$ at any point $x \in \mathcal{M}$. This *vector field* is a (generalized) *velocity* field:

question 13.2

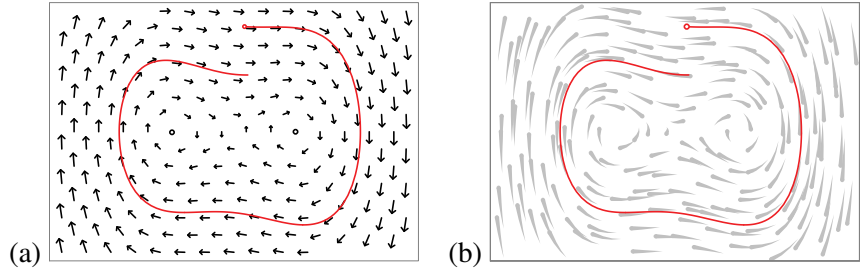
$$\dot{x}(t) = v(x). \quad (2.9)$$

Newton’s laws, Lagrange’s method, or Hamilton’s method are all familiar procedures for obtaining a set of differential equations for the vector field $v(x)$ that describes the evolution of a mechanical system. Equations of mechanics may appear different in form from (2.9), as they often involve higher time derivatives, but an equation that is second or higher order in time can always be rewritten as a set of first order equations.

We are concerned here with a much larger world of general flows, mechanical or not, all defined by a time-independent vector field (2.9). At each point of the state space, a vector indicates the local direction in which the orbit evolves. The length of the vector $|v(x)|$ is the *speed* at point x , and the direction and length of $v(x)$ changes from point to point (warning: we have slipped into a highly non-trivial notion of a “norm” or distance in state space). When the state space is a complicated manifold embedded in $\mathbb{R}^d$, one can no longer think of the vector field

remark 6.1

Figure 2.4: (a) The 2-dimensional vector field for the Duffing system (2.24), together with a short trajectory segment. (b) The flow lines. Each ‘comet’ represents the same time interval of a trajectory, starting at the tail and ending at the head. The longer the comet, the faster the flow in that region.



as being embedded in the state space. Instead, we have to imagine that each point x in this state space is attached to a different tangent plane $T\mathcal{M}_x$. The vector field lives in the union of all these tangent planes, a space called the *tangent bundle*

$$\mathbf{TM} = \bigcup_{x \in \mathcal{M}} T\mathcal{M}_x. \quad (2.10)$$

$T\mathcal{M}_x$ is called a *fiber* at x , hence the whole thing is called the *fiber bundle*. Locally a fiber bundle looks like the product of two $\mathbb{R}^d$ spaces. Just relax: we’ll do our best not to use such words again.

A simple example of a flow defined by a 2-dimensional vector field $v(x)$ is afforded by the unforced Duffing system, figure 2.4. The Lorenz flow of figure 2.5, and Rössler flow of figure 2.6, are representative 3-dimensional flows.



example 2.1
p. 62



example 2.2
p. 62



example 2.3
p. 62

The instantaneous velocity vector v is tangent to the orbit, except at the *equilibrium points*, where it vanishes.

$$\text{If } v(x_q) = 0, \quad (2.11)$$

x_q is also referred to as a *stationary, fixed, critical, invariant, rest, stagnation point, zero* of the vector field v , *standing wave, stationary solution, or steady state*. Our usage will be ‘equilibrium’ for a flow, ‘fixed point’ for a map. The orbit remains forever stuck at x_q . Otherwise the orbit passing through x_0 at time $t = 0$ can be obtained by integrating the equations (2.9):

$$x(t) = f^t(x_0) = x_0 + \int_0^t d\tau v(x(\tau)), \quad x(0) = x_0. \quad (2.12)$$

We shall consider here only *autonomous* flows, i.e., flows for which the vector field v_i is *stationary*, not explicitly dependent on time. A non-autonomous system

$$\frac{dy}{d\tau} = w(y, \tau), \quad (2.13)$$

can always be converted into a system where time does not appear explicitly. To do so, extend (‘suspend’) state space to be $(d+1)$ -dimensional by defining $x = \{y, \tau\}$, with a stationary vector field

$$v(x) = \begin{bmatrix} w(y, \tau) \\ 1 \end{bmatrix}. \quad (2.14)$$

The new flow $\dot{x} = v(x)$ is autonomous, and the orbit $y(\tau)$ can be read off $x(t)$ by ignoring the last component of x .



exercise 2.4
exercise 2.5

exercise 6.3

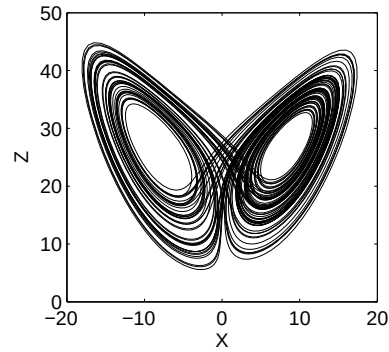


Figure 2.5: Lorenz “butterfly” strange attractor. (J. Halcrow)

Figure 2.6: A trajectory of the Rössler flow at time $t = 250$. (G. Simon)

2.2.1 Lagrangian and Eulerian viewpoints



Continuum mechanics offers two profoundly different but mathematically equivalent ways to represent a given state space flow, the ‘Lagrangian’ and the ‘Eulerian’ viewpoints. From the Eulerian perspective, one only cares about the current state of the system. Think of a field of grass, where each blade of grass represents a local velocity vector. From the Lagrangian viewpoint, one cares about where a state space point comes from, and where it is going to. Think of this state space as a bowl of linguini, where each noodle is an orbit labelled with an x_0 somewhere along its path. In the Eulerian formulation, the flow is defined by specifying (2.9), the velocity field $v(x)$. In the Lagrangian formulation, it is given by the finite time flow (2.12), i.e., the totality of the trajectories $x(t)$ comprising the deformed region, labeled by their origin x_0 in the initial undeformed region. If we mark the orbit $x(t)$ by its initial point x_0 , we are describing the flow in the *Lagrangian coordinates*. The *Eulerian* velocity $v(x)$ at a fixed state space position x is equal to the *Lagrangian* velocity $v(x(t))$ at the orbit passing through x at the instant t . Because f^t is a single-valued function, any point on the orbit can be used to label the orbit. The transport of the ‘material point’ x_0 at $t = 0$ to its value at the current point $x(t) = f^t(x_0)$ is a coordinate transformation from the Lagrangian coordinates to the Eulerian coordinates.

In numerical work we are given the equations of motion (the *local* Eulerian velocity field $v(x)$), but we care about the solutions of these equations (the *global* Lagrangian flow). Conversely, in experimental work we observe ensembles of

Lagrangian trajectories from which we then extract the velocity field (in fluid dynamics this is achieved by particle image velocimetry (PIV)). Once an Eulerian velocity field has been specified or extracted from the observational data, it is straightforward to compute the Lagrangian trajectories, objects of great practical interest in studies of long time dynamics, mixing, and transport.



fast track:
chapter 3, p. 68

2.3 Changing coordinates

Whatever else the students will need in later life, it is certain that they will have to handle changes of variables [...]. One should spend some time teaching in wealth of detail relevant changes of variables. Luckily, some of these are still included in textbooks, though no textbook now in print awards this essential technique the importance it deserves. Worse, no one realizes that changes of variables are not just a trick; they are a coherent theory [...].

— *Ten Lessons*, by Gian-Carlo Rota [39]

Problems are handed down to us in many shapes and forms, and they are not always expressed in the most convenient way. In order to simplify a given problem, one may stretch, rotate, bend and mix the coordinates, but in doing so, the velocity vector field will also change. The vector field lives in a (hyper)plane tangent to the state space (remember the dreaded tangent bundle?), so changing the state space coordinates affects the coordinates of the tangent space as well, in a way that we will now describe.



Denote by h the *conjugation function* which maps the coordinates of the initial state space $\mathcal{M}$ into the reparameterized state space $\tilde{\mathcal{M}} = h(\mathcal{M})$, with a point $x \in \mathcal{M}$ related to a point $y \in \tilde{\mathcal{M}}$ by

$$y = h(x) = (y_1(x), y_2(x), \dots, y_d(x)).$$

The change of coordinates must be one-to-one, a diffeomorphism on open neighborhoods in $\mathcal{M}$ and $\tilde{\mathcal{M}}$, so given any point y we can go back to $x = h^{-1}(y)$. For smooth flows the reparameterized dynamics should support the same number of derivatives as the initial one. If h is a (piecewise) analytic function, we refer to h as a *smooth conjugacy*.

What form does the velocity vector field $\dot{x} = v(x)$ take in the new coordinate system $y = h(x)$? Let's compute it first for a 1-dimensional dynamical system. Let $x(t) = f^t(x)$ be the solution to the differential equation $\dot{x} = v(x)$ starting at x , and $y(t) = g^t(y)$ be the solution to the same problem, but in the new coordinates. The velocity vector field in the new coordinates follows from the chain rule:

$$w(y) = \left. \frac{dg^t}{dt}(y) \right|_{t=0} = \frac{dy}{dx} \frac{dx}{dt} = \frac{dh}{dx} v(x).$$

To understand this transformation for a state space of arbitrary dimension, a little geometrical intuition pays off. The evolution rule $g^t(y_0)$ on $\tilde{\mathcal{M}}$ can be computed from the evolution rule $f^t(x_0)$ on $\mathcal{M}$ by taking the initial point $y_0 \in \tilde{\mathcal{M}}$, going back to $\mathcal{M}$, evolving, and then mapping the final point $x(t)$ back to $\tilde{\mathcal{M}}$:

$$y(t) = g^t(y_0) = h \circ f^t \circ h^{-1}(y_0). \quad (2.15)$$

Here ‘ $\circ$ ’ stands for functional composition $h \circ f(x) = h(f(x))$, so (2.15) is a shorthand for $y(t) = h(f^t(h^{-1}(y_0)))$. Hence, $h(x)$ is called a ‘conjugating function’; it is a similarity transformation generalized to nonlinear coordinate transformations.

The vector field $\dot{x} = v(x)$ is locally tangent to the flow f^t ; it is related to the flow by differentiation (2.8) along the orbit. The vector field $\dot{y} = w(y)$, $y \in \tilde{\mathcal{M}}$ locally tangent to g^t , follows by the chain rule:

exercise 2.10

$$\begin{aligned} w(y) &= \left. \frac{dg^t}{dt}(y) \right|_{t=0} = \left. \frac{d}{dt} (h \circ f^t \circ h^{-1}(y)) \right|_{t=0} \\ &= h'(h^{-1}(y)) v(h^{-1}(y)) = h'(x) v(x). \end{aligned} \quad (2.16)$$

In order to rewrite the right-hand side as a function of y , note that the ∂_y differentiation of $h(h^{-1}(y)) = y$ implies

$$\left. \frac{\partial h}{\partial x} \right|_x \cdot \left. \frac{\partial h^{-1}}{\partial y} \right|_y = 1 \quad \rightarrow \quad \frac{\partial h}{\partial x}(x) = \left[\frac{\partial h^{-1}}{\partial y}(y) \right]^{-1}. \quad (2.17)$$

Thus, the equations of motion in the transformed coordinates, with the indices reinstated, are

$$\dot{y}_i = w_i(y) = \left[\frac{\partial h^{-1}}{\partial y}(y) \right]_{ij}^{-1} v_j(h^{-1}(y)). \quad (2.18)$$

Imagine the state space as a rubber sheet with the flow lines drawn on it. A coordinate change h corresponds to pulling and tugging on the rubber sheet smoothly, without cutting, gluing, or self-intersections of the distorted rubber sheet. Trajectories that are closed loops in $\mathcal{M}$ will remain closed loops in the new manifold $\tilde{\mathcal{M}}$, but their shapes will change. Globally, h deforms the rubber sheet in a highly nonlinear manner, but locally it simply rescales and shears the tangent field by the coordinate transformation Jacobian matrix $\partial_j h_i$, yielding the simple transformation law (2.16) for the velocity fields.

Time itself is a parametrization of points along flow lines, and it can also be reparameterized, $s = s(t)$, with the concomitant modification of (2.18). An example is the 2-body collision regularization of the helium Hamiltonian (8.32), to be undertaken in appendix A2.2.

appendix A2.2

In chapter 30 we shall dispose of the fear of ‘infinite-dimensional’ dynamical systems—you might prefer to skip sect. 2.4 on first reading.

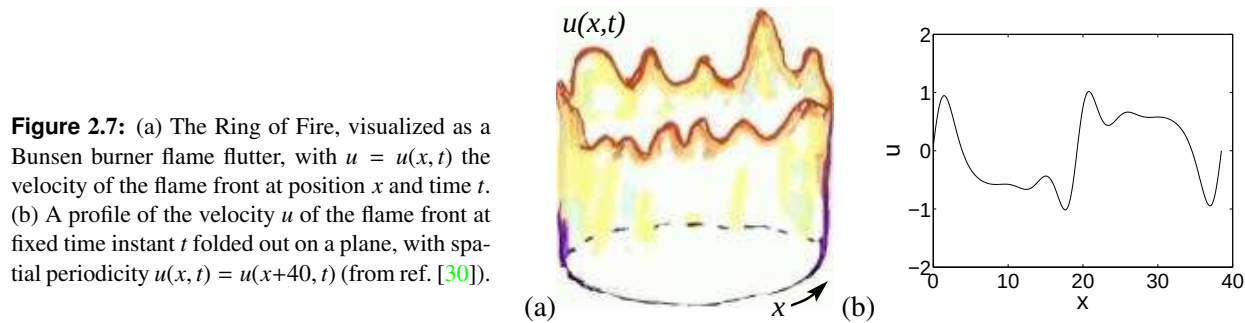


Figure 2.7: (a) The Ring of Fire, visualized as a Bunsen burner flame flutter, with $u = u(x, t)$ the velocity of the flame front at position x and time t . (b) A profile of the velocity u of the flame front at fixed time instant t folded out on a plane, with spatial periodicity $u(x, t) = u(x+40, t)$ (from ref. [30]).

2.4 Life in extreme dimensions

Sometimes I've believed as many as six impossible things before breakfast.

— Lewis Carroll



Systems described by partial differential equations [PDEs] are said to be ‘infinite dimensional’ dynamical systems, because in order to uniquely specify the state of a spatially extended ‘field’, one needs infinitely many numbers, one for the value of the field at each configuration space point. Even though the state space is infinite-dimensional, the long-time dynamics of many such systems of physical interest is finite-dimensional, contained within a ‘strange attractor’ or an ‘inertial manifold’. Most of us find it hard to peer into four dimensions. How are we to visualize -and why we would have any hope of visualizing- dynamics in such extreme dimensions? A representative point is a point, and its trajectory is a curve in any 2- or 3-dimensional projection, so that is not so hard. What is hard is to get an understanding of relative disposition of different states. The coordinates have to be chosen thoughtfully, since most of the interesting orbits appear minuscule in a randomly picked coordinate frame.

chapter 30

A dynamical system is specified by the pair $(\mathcal{M}, f)$, where d numbers uniquely determine a state of the system, or the representative point x in the state space manifold $\mathcal{M}$. Here we focus on how one constructs such state space, and how one visualizes a representative point x and its trajectory $f^t(x)$ time t later. We shall return to dynamics in chapter 30, where we examine the evolution rule f^t that maps a state space region $\mathcal{M}_i$ into the region $f^t(\mathcal{M}_i)$ (see figure 2.2). Also in chapter 30, we discuss the time-evolution equations for spatially-extended systems and the ‘turbulence’ that such systems exhibit.

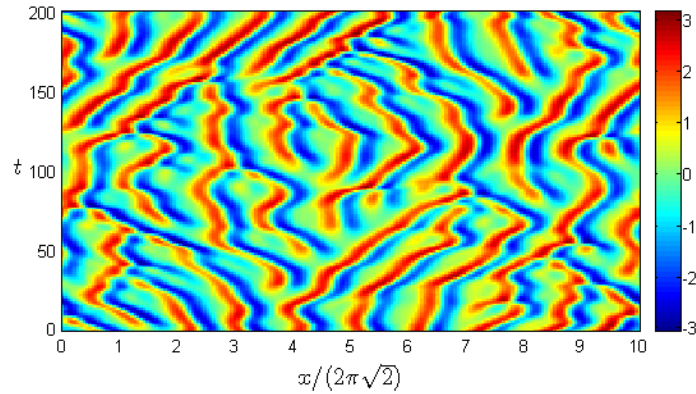
chapter 30

2.4.1 Configuration space: a fluttering flame front



Consider the flame front flutter of gas burning on your kitchen stove. Such a ‘Bunsen burner’, invented by Göttingen chemistry prodigy Robert Bunsen in 1855, entered popular culture in 1963 with the song “Ring of Fire” by Johnny Cash *et al.* [6]. The flame front instabilities of a ‘Bunsen burner’ are perhaps the most familiar example of a nonlinear system that exhibits ‘turbulence’ (or, more modestly, ‘spatiotemporally chaotic behavior’): a typical configuration space (or

Figure 2.8: A spatiotemporal plot of the Ring of Fire “turbulent” solution, periodic domain $u(x, t) = u(x + 20\pi\sqrt{2}, t)$ is obtained by plotting the profile of figure 2.7 (b) for successive time instants (vertical axes). The color indicates the value of u at a given position and instant in time (from ref. [9]).



the much abused word ‘physical’ space) visualization is sketched in figure 2.7. Its state can be described by the ‘flame front velocity’ $u = u(x, t)$ on a periodic domain $u(x, t) = u(x + L, t)$.

The spatial, ‘configuration’ or ‘physical’ state space visualization of systems, such as the ‘Ring of Fire’ (figure 2.7), offers little insight into their detailed *dynamics*. Furthermore, neither we do gain any additional insight into these dynamics by studying a fixed time snapshot of velocity and vorticity fields using 3D Navier-Stokes equations, a graphical representation of the flame front flutter in time (figure 2.8), nor even a [time-evolving video](#) of a fluid. To truly understand the dynamics of such systems, one must turn to the complementary, and often more illuminating *state space* representations. In this context ‘flow’ refers to a d -dimensional flow in the dynamical state space, not the flow of a fluid, and ‘velocity’ to the state space tangent field $\dot{x} = v(x)$, not to the 3D configuration space fluid velocity field $u(x, t) \in \mathbb{R}^3$. A ‘representative point’ is a full specification of the state $x \in \mathcal{M}$ of the system. In today’s experiments or numerical simulations, this is a set of anything from 16 to 10^6 numbers needed to specify a complete snapshot of the flame front figure 2.7, or the state of volume of turbulent fluid in a pipe at an instant in time.

[chapter 30](#)

2.4.2 Constructing a state space

Think globally, act locally.

— Patrick Geddes



At this juncture, our everyday, plumber’s visual intuition actually interferes with dynamical visualization of state space of a spatially-extended systems: while the spatial dimension of the Ring of Fire is 1, its dimension as a dynamical system is ∞ . Absorbing this simple fact of life is the same rite of passage as going from the 1-degree of freedom quantum mechanical oscillator to the ‘second quantization’ of quantum field theory, with its infinitely many quantum oscillator degrees of freedom.

To develop some intuition about such dynamics we turn to experiments, or numerical simulations, such as the Ring of Fire time evolution, figure 2.8. The first

thing we note is that while the dynamics might be ‘turbulent’, for many such systems the long-time solutions tend to be smooth. This suggests that a discretization, perhaps aided by interpolations such as n -point spatial derivatives, might give us a representation of the dynamics of reasonable accuracy.

Discrete mesh: You can subdivide the configuration domain into a sufficiently fine discrete grid of N boxes, replace space derivatives in the governing equations by approximate discrete derivatives, and integrate a finite set of first order differential equations for the discretized spatial components $u_j(t) = u(jL/N, t)$, by any integration routine you trust. Most often that’s the best you can do.

The next thing we note is that the solutions for many physical systems of physical interest tend to be not only smooth, but also that the laws that govern them are invariant in form under operations such as translations. For example, in configuration space the fluttering flame front governing equations should be invariant in their form under rotations, time translations, and reflection $x \rightarrow -x$, $u \rightarrow -u$.

Spectral methods: The spatial periodicity $u(x, t) = u(x + L, t)$ then suggests that it might be convenient to work in the Fourier space,

$$u(x, t) = \sum_{k=-\infty}^{+\infty} \tilde{u}_k(t) e^{iq_k x}, \quad (2.19)$$

where $\tilde{u}_k = x_k + i y_k = |\tilde{u}_k| e^{i\phi_k}$, $q_k = 2\pi k/L$, L is the domain size, x is the spatial coordinate and t is time. Thus a state of a spatially 1-dimensional extended system can be described by an infinite set of complex Fourier coefficients $\tilde{u}_k(t)$. The velocity field $u(x, t)$ is real, so $\tilde{u}_k = \tilde{u}_{-k}^*$, and we can replace the sum by an $k \geq 0$ sum, with u written as its reflection-symmetric part (sum of cosines) plus its reflection-antisymmetric part (sum of sines). This is an example of an infinite-dimensional *state space*, which we alluded to on page 50.

example 12.7

Intuitively the flame front is smooth, so Fourier coefficients $\tilde{u}_k$ drop off fast with k , and truncations of (2.19) to finite numbers of terms can yield highly accurate states. In numerical computations this state space is truncated to a finite number of real numbers. For example, a state might be specified by $2N$ real Fourier coefficients, or ‘computational degrees of freedom’

$$x = (x_1, y_1, x_2, y_2, \dots, x_N, y_N)^T. \quad (2.20)$$

More sophisticated variants of such truncations are called in the literature *Galerkin truncations*, or *Galerkin projections*.

Once a trajectory is computed in Fourier space, we can recover and plot the corresponding spatiotemporal pattern $u(x, t)$ over the configuration space, as in figure 2.7 and figure 2.8, by inverting (2.19). Spatiotemporal patterns give us a qualitative picture of the flow and a physical intuition about the energetics of the flow, but no detailed dynamical information; for that, tracking the evolution in a high-dimensional state space is much more informative.

2.4.3 State space, as visualized by dummies

This is dedicated to Comrade Student Xiong Ding
 — Prof. Dr. Dr. h.c. Gatto Nero, Grande Ufficiale
 dell'Ordine della Stella della Notorietà Italiana



So the simplest way to construct (in practice a finite dimensional approximation to) state space coordinates is by a discrete mesh $u(x, t) \rightarrow u_j(t)$ or ‘spectral’ coefficients $u(x, t) \rightarrow \tilde{u}_k(t)$. We shall refer to such coordinates as ‘computational degrees of freedom’. The same dynamics can look *very* different in different choices of coordinates. And when we say that the dynamics is ‘61,506-dimensional’, we mean that in order to capture a particular physical observable to a sufficient number of digits of accuracy, we need at least 61,506 computational degrees of freedom.

The question is: how is one to look at such state space flow? The laziest thing to do is to examine the trajectory’s projections onto any three computational degrees of freedom, let’s say the first three Fourier modes $(\tilde{u}_1, \tilde{u}_2, \tilde{u}_3)$. Why would you do that? Well, that’s what a computer spews out. This won’t do. Let’s accept that you do not know much about high dimensions, but you have been born someplace where they force you to watch athletes kick a ball, for hours on end. Your choice of $(\tilde{u}_1, \tilde{u}_2, \tilde{u}_3)$ coordinates means that you (or the TV camera) are standing at a corner of the field. Far, far away, at the opposite end of the field, there is action - but you only see a few little moving silhouettes, and can hardly see the ball.

Or, if you are more inclined towards scholarly activities and would rather whittle away hours away evaluating Meijer G -functions, here is a precise way of saying the same thing: choose a direction in a high-dimensional state space and call it basis vector $\mathbf{e}^{(1)}$. Now randomly pick a state $\mathbf{u}$ in state space, and this gives you have a second vector. What is the angle between these two vectors? The cosine of that angle you compute by evaluating the ‘dot’ product (or L2 norm)

$$\langle \mathbf{u} | \mathbf{e}^{(1)} \rangle = \frac{1}{V} \int_{\Omega} dx \, \mathbf{u} \cdot \mathbf{e}^{(1)}, \quad \|\mathbf{u}\|^2 = \langle \mathbf{u} | \mathbf{u} \rangle. \quad (2.21)$$

Once you finish the exercise 2.13, you will know what every computer scientist knows: the expectation value of the angle between any two high-dimensional vectors picked at random is 90° , with a very small variance. In other words, when using a random coordinate system in high dimensions, every distant silhouette of Cristiano Ronaldo is vanishingly small. And as your lazy $(\tilde{u}_1, \tilde{u}_2, \tilde{u}_3)$ coordinates are a random choice, your turbulent state might require 10^5 such coordinates to be accurately resolved.

exercise 2.13

So, if you were a referee, or a camera operator, would you really just stand there, in the far corner of the field?

2.4.4 Exact state-space portraiture: go where the action is

(J.F. Gibson and P. Cvitanović)

Let's say you are interested in dynamics, especially recurrent dynamics, so you scan the soccer field to identify prominent states of interest to you. You then choose to characterize and observe recurrent coherent structures by using long-time numerical simulations. If you form a basis set from these structures and project the evolving state $x(t)$ onto this basis, coordinates so formed will capture close recurrences to these states. That is, you form orthonormal basis functions $\{\mathbf{e}^{(1)}, \mathbf{e}^{(2)}, \dots, \mathbf{e}^{(n)}\}$ from a set of linearly independent fluid states and produce a state-space trajectory

$$x(t) = (x_1(t), x_2(t), \dots, x_n(t), \dots), \quad x_n(t) = \langle \mathbf{u}(t) | \mathbf{e}^{(n)} \rangle \quad (2.22)$$

in the $\{\mathbf{e}^{(n)}\}$ coordinate frame. The *projection* of the trajectory can be viewed by a human in any of the $2d$ planes $\{\mathbf{e}^{(m)}, \mathbf{e}^{(n)}\}$ or in $3d$ perspective views $\{\mathbf{e}^{(l)}, \mathbf{e}^{(m)}, \mathbf{e}^{(n)}\}$. The dimensionality is lower than the full state space, so in such projections trajectories can appear to cross. It is important to understand that this is a low-dimensional *visualization*, not low-dimensional *modeling*, a truncation to fewer computational degrees of freedom. The dynamics are computed with fully-resolved direct numerical simulations and then projected onto basis sets to produce low-dimensional state-space portraits, tailored to specific purposes and specific regions of state space. The resulting portraiture depends on the *physical states* involved and not on the (arbitrary) choice of a numerical representation. Such well-chosen portraits reveal dynamical information visually, providing insight into dynamics that can guide further analysis.

At first glance, turbulent dynamics visualized in state space might appear hopelessly complex, but many detailed studies suggest it might be less complex than feared: turbulent dynamics appears to be pieced together from near visitations to exact invariant solutions, interspersed by transient interludes. Equilibria, traveling waves, and periodic solutions embody Hopf's vision: a repertoire of recurrent spatio-temporal patterns explored by turbulent dynamics. We conceive of turbulence as a walk through a repertoire of unstable recurrent patterns. As a turbulent flow evolves, every so often we catch a glimpse of a familiar pattern. For any finite spatial resolution, the flow approximately follows for a finite time a pattern belonging to a finite alphabet of admissible fluid states, represented in ChaosBook by a set of exact invariant solutions.

appendix A1.5

remark 30.1

There is an infinity of possible basis sets, but two types of bases appear particularly natural: (a) a global basis, determined by a set of dynamically important states, or (b) a local basis, defined, for example, in terms of a given equilibrium and its linear stability eigenvectors.

section 4.8

With this road map in hand, we can take a stroll through the state space of a spatiotemporally turbulent flow. Like many dynamical narratives, this might turn into a long trek through unfamiliar landscapes with many landmarks of local interest. It is amazing that such a promenade is possible even in 10^5 dimensions

(computational degrees of freedom). But a detailed road map is a necessary prerequisite for solving at least three of your outstanding problems: (a) uncover interrelations between (in principle infinite number of) unstable invariant solutions of a turbulent flow, (b) partition state space (symbolic dynamics) in order to systematically explore turbulent dynamics, and (c) identify linearly stable eigenvectors and their unstable-manifold continuations in order to transform a given spatiotemporal state into a desired target state.



In summary, when dealing with spatiotemporally extended systems, you'll need dual vision - you will have to think both in *configuration space* and in the *state space*.

2.5 Computing trajectories

On two occasions I have been asked [by members of Parliament], 'Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?' I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question.

— Charles Babbage



You have not learned dynamics unless you know how to integrate numerically whatever dynamical equations you face. Sooner or later, you need to implement some finite time-step prescription for integration of the equations of motion (2.9). The simplest is the Euler integrator, which advances the trajectory by adding a small vector $\delta\tau \times \text{velocity}$ at each time step:



$$x_i \rightarrow x_i + v_i(x) \delta\tau. \quad (2.23)$$

This might suffice to get you started, but as soon as you need higher numerical accuracy, you will need something better. There are many excellent reference texts and computer programs that can help you learn how to solve differential equations numerically using sophisticated numerical tools, such as pseudo-spectral methods or implicit methods. If you are interested in Hamiltonian flows you might want to implement a symplectic integrator of the type discussed in appendix A31.2.1. If a 'sophisticated' integration routine takes days and gobbles up terabits of memory, you are using brain-damaged high level software. Try writing a few lines of your own Runge-Kutta code in some mundane everyday language. While you absolutely need to master the requisite numerical methods, this is neither the time nor the place to expound upon them; how you learn them is your business. And if you have developed some nice routines for solving problems in this text or can point another student to some, let us know.

exercise 2.6

exercise 2.7

exercise 2.9

exercise 2.11

Résumé



Start from a state space point and evolve it for a finite time. Then trace out its *trajectory*. If you evolve it forward and backward for infinite time, you get its

orbit, the set of all states reachable by evolution from a given state space point. An orbit is a time-invariant notion. Time evolution marches points along its path, but the set itself does not change. The *flow* describes the time evolution of all state space points, i.e., the totality of all orbits; the evolution law f turns the state space into a bowl of spaghetti, where each individual strand of spaghetti is an orbit.

Chaotic dynamics with a low-dimensional attractor can be visualized as a succession of nearly periodic but unstable motions. In the same spirit, turbulence in spatially extended systems can be described in terms of recurrent spatiotemporal patterns. Pictorially, dynamics drives a given spatially extended system through a repertoire of unstable patterns; as we watch a turbulent system evolve, every so often we catch a glimpse of a familiar pattern. For any finite spatial resolution and finite time, the system follows approximately a pattern belonging to a finite repertoire of possible patterns. The long-term dynamics can be thought of as a walk through the space of such patterns. Recasting this image into mathematics is the subject of this book. 

State-space portraits are *dynamically intrinsic*, since their projections are defined in terms of solutions of the equations of motion, and *representation independent*, since the L2 product (2.21) is independent of the numerical representation. The method can be applied to any high-number of computational degrees of freedom discretization of a dissipative flow. Production of state-space portraits requires numerical data of configuration space fields evolving in time (obtained from simulation or experiment), estimates of important physical states (such as equilibria and their linear stability eigenfunctions), and a method of computing the inner product between velocity fields over the physical domain.

Commentary

Remark 2.1. ‘State space’ or ‘phase space?’ In ChaosBook, *state space* is the set of admissible states in a general d - or ∞ -dimensional dynamical system. The term *phase space* is reserved for Hamiltonian state spaces of $2D$ -dimensions, where D is the number of Hamiltonian degrees of freedom. If the state space is a continuous smooth manifold much of the literature [25, 32] refers to it as ‘phase space,’ but we find the control engineering usage sharper: in the state space (or ‘time-domain’) description of an autonomous physical system, the instantaneous state of the system is represented as a point within the ‘state space,’ a space whose axes are the state variables, and the evolution of a state is given by differential equations which are first-order in time. Hopf [26] would refer to such a state as an ‘instantaneous phase’ of the system obeying a ‘differential law of the phase motion’. The distinction made here is needed in a text where one treats deterministic dynamical systems, stochastic systems and quantum-mechanical systems on equal footing. The term ‘phase’ has a precise meaning in wave mechanics, quantum mechanics and dynamics of integrable systems at the heart of Hamilton’s formulation of Newtonian mechanics, while ‘state space’ is more descriptive of the way the notion is used in the general theory of dynamical systems. Further confusion arises when prefix spatio- as in ‘spatiotemporal’ is used in reference to states extended in the (1, 2, or 3-dimensional) physical configuration space. They may exhibit spatial wave-like behaviors, but their *state space* is ∞ -dimensional.

Much of the literature denotes the vector field in a first order differential equation (2.9) by $f(x)$ or $F(x)$ or even $X(x)$, and its integral for time t by the ‘time- t forward map’ or ‘flow map’ $x(x_0, t) = \Phi(x_0, t)$, or $\phi_t(x_0)$, or something else. Here we treat maps and flows on an equal footing, and we save Greek letters for matters quantum-mechanical. We reserve the notation $f^t(x)$ for *maps* such as (2.12) and refer to a state space velocity *vector field* as $v(x)$. We come to regret this choice very far into the text, only by the time we delve into Navier-Stokes equations.

Question 2.1. Alice’s professor begs to differ.

Q Why do you call these things ‘periodic orbits’? My professor refers to them as ‘non-chaotic solutions’.

A It’s an insult to Henri Poincaré. He called periodic orbits ‘periodic orbits’, so why rename them? It’s like calling $3/4$ a “nonreal” number. The way one measures continuum is by bracketing a real number by increasingly refined intervals, each finitely specified, let’s say by a rational number $3/4$. The way ergodic theory works is that the state space is partitioned (coarse grained) into recurrent subregions, each containing within it an exactly recurrent point (or a compact invariant subset, such as Poincaré’s relative periodic orbit around L2, etc., etc.), which one can use as a marker. You can name it after your Russian aunt’s *hoola hoop*, but that is not yet widely used nomenclature.

Remark 2.2. Rössler and Duffing flows. The Duffing system (2.24) arises in the study of electronic circuits [15]. The Rössler flow (2.30) is the simplest flow which exhibits many of the key aspects of chaotic dynamics. It was introduced in ref. [38] as a set of equations describing no particular physical system, but capturing the essence of Lorenz chaos in the most simple of smooth flows. Otto Rössler, a man of classical education, was inspired in this quest by that rarely cited grandfather of chaos, Anaxagoras (456 B.C.). This and references to earlier work can be found in refs. [22, 36, 46]. We recommend in particular the inimitable Abraham and Shaw illustrated classic [1] for its beautiful sketches of many flows, including the Rössler flow. Timothy Jones [28] has a number of interesting simulations on a Drexel website.

example 2.3

The Rössler flow is the simplest flow which exhibits many of the key aspects of chaotic dynamics; we shall use it and the 3-pinball systems throughout ChaosBook to motivate introduction of Poincaré sections, return maps, symbolic dynamics, cycle expansions, and much else. Rössler flow is integrated in exercise 2.7, its equilibria are determined in exercise 2.8, its Poincaré sections constructed in exercise 3.1, and the corresponding return map computed in exercise 3.2. Its volume contraction rate is computed in exercise 4.3, its topology investigated in exercise 4.4, the shortest Rössler flow cycles are computed and tabulated in exercise 7.1, and its Lyapunov exponents evaluated in exercise 6.4.

chapter 9

Remark 2.3. Lorenz equation. The Lorenz equation (2.25) is the most celebrated early illustration of “deterministic chaos” [32] (but not the first - that honor goes to Dame Cartwright [5] in 1945. Amusingly, Denisov and Ponomarev [11] argue that Ben F. Laposky might have been the first to observe chaotic attractors as early as 1953, which, strictly speaking falls after 1945, even in Russia). Lorenz’s 1963 paper, which can be found in reprint collections refs. [8, 24], is a pleasure to read, and it is still one of the best introductions to the physics motivating such models (read more about Lorenz [here](#)). The equations, a set of ODEs in $\mathbb{R}^3$, exhibit strange attractors. W. Tucker [47–49] has proven rigorously (via interval arithmetic) that the Lorenz attractor is strange for the original parameters (no stable orbits) and that it has a long stable periodic orbit for slightly different parameters. In contrast to the hyperbolic strange attractors such as the weakly

perturbed cat map [7], the Lorenz attractor is structurally unstable. Frøyland [17] has a nice brief discussion of Lorenz flow. Frøyland and Alfsen [18] plot many periodic and heteroclinic orbits of the Lorenz flow; some of the symmetric ones are included in ref. [17]. Galias and Tucker [19] compute all 2536 periodic orbits of symbolic dynamics periods $n \leq 14$. Guckenheimer-Williams [23] and Afraimovich-Bykov-Shilnikov [2] offer an in-depth discussion of the Lorenz equation. The most detailed study of the Lorenz equation was undertaken by Sparrow [43]. For a geophysics derivation, see Rothman course notes [40]. For a physical interpretation of ρ as “Rayleigh number,” see Jackson [27] and Seydel [41]. The Lorenz truncation to 3 modes, however, is so drastic that the model bears no relation to the geophysical hydrodynamics problem that motivated it. Just for fun, as Lorenz was such a lovable weatherman, in 1972 Willem Malkus constructed [33], by a feat of reverse engineering, a physical system, a “water wheel”, popularized by Strogatz [45], that is described by Lorenz equations. You can see it simulated [Wolfram.com](#), and tested experimentally at <http://www.ace.gatech.edu>. There is no deep physics in this lovely game, it is but a cute distraction. For detailed pictures of Lorenz invariant manifolds consult Vol II of Jackson [27] and “Realtime visualization of invariant manifolds” by Ronzan. The Lorenz attractor is a very thin fractal – as we shall see, stable manifold thickness is of the order 10^{-4} – whose fractal structure has been accurately resolved by D. Viswanath [50, 51]. If you wonder what analytic function theory has to say about Lorenz, check ref. [52]. Modular flows are your thing? E. Ghys and J. Leys have a beautiful tale for you. Refs. [31, 34] might also be of interest. (continued in remark 11.1)

Remark 2.4. High-dimensional flows and their visualizations. Dynamicist’s vision of turbulence was formulated by Eberhard Hopf in his seminal 1948 paper [25], see appendix A1.5. Computational neuroscience grapples with closely related visualization and modeling issues [16, 20]. Much about high-dimensional state spaces is counterintuitive. The literature that discusses the 90° angle found between any two high-dimensional vectors is mostly about spikey spheres: see the draft of the Hopcroft and Kannan [4] book and Ravi Kannan’s course; lecture notes by Hermann Flaschka on *Some geometry in high-dimensional spaces*; Wegman and Solka [53] visualizations of high-dimensional data; Spruill paper [44]; a lively mathoverflow.org thread on “Intuitive crutches for higher dimensional thinking.”

appendix A1.5

exercise 2.13

The ‘good’ coordinates, introduced in ref. [21] and described here are akin in spirit to the low-dimensional projections of proper orthogonal decomposition (POD) modeling [3], in that both methods aim to capture key features and dynamics of the system in just a few dimensions. But the method described here is very different from POD in a key way: we construct basis sets from *exact solutions of the fully-resolved dynamics* rather than from the empirical eigenfunctions of the POD. These exact solutions and their linear stability modes (a) characterize the spatially-extended states precisely, as opposed to the truncated expansions of the POD, (b) allow for different basis sets and projections for different purposes and different regions of state space, (c) visualize every point in these projections as a point in the full state space, but do not suggest low-dimensional ODE models, and (d) are not limited to Fourier mode bases.

(J.F. Gibson and P. Cvitanović)

Remark 2.5. Dynamical systems software: First of all, to understand how adults in the room feel about this matter, consult Gian-Carlo Rota [39]. But integrate we must, so: J.D. Meiss [35] has maintained for many years *Sci.nonlinear FAQ* which is now in part superseded by the SIAM Dynamical Systems website www.dynamicalsystems.org. The website glossary contains most of Meiss’s FAQ plus new ones, as well as an up-to-date software list [42] with links to DSTool, xpp, AUTO [13, 14], etc.. Kuznetsov [29]

Appendix D.9 gives an exhaustive overview of software available in 2004. More recent are *E&F Chaos* of Diks *et al.* [12] and Datseris [10] *DynamicalSystems.jl* Julia package. For further links to online codes check ChaosBook.org/extras, as well as remark 15.1.

References

- [1] R. H. Abraham and C. D. Shaw, *Dynamics - The Geometry of Behavior* (Wesley, Reading, MA, 1992).
- [2] V. S. Afraimovich, B. B. Bykov, and L. P. Shilnikov, “On the origin and structure of the Lorenz attractor”, *Dokl. Akad. Nauk SSSR* **234**, In Russian, 336–339 (1977).
- [3] N. Aubry, P. Holmes, J. L. Lumley, and E. Stone, “The dynamics of coherent structures in the wall region of turbulent boundary layer”, *J. Fluid Mech.* **192**, 115–173 (1988).
- [4] A. Blum, J. Hopcroft, and R. Kannan, *Foundations of Data Science* (Cambridge Univ. Press, Cambridge UK, 2020).
- [5] M. L. Cartwright and J. E. Littlewood, “On non-linear differential equations of the second order”, *J. London Math. Soc.* **20**, 180–189 (1945).
- [6] J. Cash, J. Carter Cash, and M. Kilgore, *The Ring of Fire*, New York, 1963.
- [7] S. C. Creagh, “Quantum zeta function for perturbed cat maps”, *Chaos* **5**, 477–493 (1995).
- [8] P. Cvitanović, *Universality in Chaos*, 2nd ed. (Adam Hilger, Bristol, 1989).
- [9] P. Cvitanović, R. L. Davidchack, and E. Siminos, “On the state space geometry of the Kuramoto-Sivashinsky flow in a periodic domain”, *SIAM J. Appl. Dyn. Syst.* **9**, 1–33 (2010).
- [10] G. Datseris, “DynamicalSystems.jl: A Julia software library for chaos and nonlinear dynamics”, *J. Open Source Soft.* **3**, 598 (2018).
- [11] S. Denisov and A. V. Ponomarev, “Oscillons: An encounter with dynamical chaos in 1953?”, *Chaos* **21**, 023123 (2011).
- [12] C. Diks, C. Hommes, V. Panchenko, and R. van der Weide, “E&F Chaos: A user friendly software package for nonlinear economic dynamics”, *Computational Econ.* **32**, 221–244 (2008).
- [13] E. J. Doedel, *Lecture Notes on Numerical Analysis of Nonlinear Equations*, tech. rep. (Concordia Univ., 2010).
- [14] E. J. Doedel, A. R. Champneys, T. F. Fairgrieve, Y. A. Kuznetsov, B. Sandstede, and X. Wang, *AUTO: Continuation and Bifurcation Software for Ordinary Differential Equations*, tech. rep. (Computational Mathematics and Visualization Laboratory, Concordia Univ., 2007).
- [15] G. Duffing, *Erzwungene Schwingungen bei veränderlicher Eigenfrequenz und ihre technische Bedeutung*, 41–42 (Vieweg und Sohn, Braunschweig, 1918).

- [16] A. Farshchian, J. A. Gallego, J. P. Cohen, Y. Bengio, L. E. Miller, and S. A. Solla, [Adversarial domain adaptation for stable brain-machine interfaces](#), in [International Conference on Learning Representations](#) (2019), pp. 1–14.
- [17] J. Frøyland, [Introduction to Chaos and Coherence](#) (Taylor & Francis, Bristol, 1992).
- [18] J. Frøyland and K. H. Alfsen, “Lyapunov-exponent spectra for the Lorenz model”, [Phys. Rev. A](#) **29**, 2928 (1984).
- [19] Z. Galias and W. Tucker, [Symbolic dynamics based method for rigorous study of the existence of short cycles for chaotic systems](#), in [2009 IEEE intern. symp. circuits and systems](#) (2009), pp. 1907–1910.
- [20] J. A. Gallego, M. G. Perich, R. H. Chowdhury, S. A. Solla, and L. E. Miller, “Long-term stability of cortical population dynamics underlying consistent behavior”, [Nature Neuroscience](#) **23**, 260–270 (2020).
- [21] J. F. Gibson, J. Halcrow, and P. Cvitanović, “Visualizing the geometry of state-space in plane Couette flow”, [J. Fluid Mech.](#) **611**, 107–130 (2008).
- [22] R. Gilmore and M. Lefranc, [The Topology of Chaos](#), 2nd ed. (Wiley, New York, 2003).
- [23] J. Guckenheimer and R. Williams, “Structural stability of the Lorenz attractor”, [Publ. Math. IHES](#) **50**, 55–72 (1979).
- [24] B.-L. Hao, [Chaos II](#) (World Scientific, Singapore, 1990).
- [25] E. Hopf, “A mathematical example displaying features of turbulence”, [Commun. Pure Appl. Math.](#) **1**, 303–322 (1948).
- [26] E. Hopf, “Statistical hydromechanics and functional calculus”, [J. Rat. Mech. Anal.](#) **1**, 87–123 (1952).
- [27] E. A. Jackson, [Perspectives of Nonlinear Dynamics](#), Vol. 1 (Cambridge Univ. Press, Cambridge UK, 1989).
- [28] T. Jones, [Symmetry of Chaos Animations](#), tech. rep. (Nonlinear Dynamics Group, Drexel Univ., 2012).
- [29] Y. A. Kuznetsov, [Elements of Applied Bifurcation Theory](#), 3rd ed. (Springer, New York, 2004).
- [30] Y. Lan and P. Cvitanović, “Unstable recurrent patterns in Kuramoto-Sivashinsky dynamics”, [Phys. Rev. E](#) **78**, 026208 (2008).
- [31] J. B. Laughlin and P. C. Martin, “Transition to turbulence of a statically stressed fluid”, [Phys. Rev. Lett.](#) **33**, 1189 (1974).
- [32] E. N. Lorenz, “Deterministic nonperiodic flow”, [J. Atmos. Sci.](#) **20**, 130–141 (1963).
- [33] W. V. R. Malkus, “Non-periodic convection at high and low Prandtl number”, [Mem. Societe Royale des Sciences de Liege](#), 125–128 (1972).
- [34] P. Manneville and Y. Pomeau, “Different ways to turbulence in dissipative dynamical systems”, [Physica D](#) **1**, 219–226 (1980).
- [35] J. D. Meiss, [Sci.nonlinear FAQ, Computational Resources](#), 2003.

- [36] J. Peinke, J. Parisi, O. E. Rössler, and R. Stoop, *Encounter with Chaos - Self Organised Hierarchical Complexity in Semiconductor Experiments* (Springer, New York, 1992).
- [37] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in Fortran 77*, 2nd ed. (Cambridge Univ. Press, Cambridge UK, 1996).
- [38] O. E. Rössler, “An equation for continuous chaos”, *Phys. Lett. A* **57**, 397 (1976).
- [39] G.-C. Rota, *Ten lessons I wish I had learned before I started teaching differential equations*, tech. rep., Simmons College Meeting of Math. Assoc. of America (Massachusetts Inst. of Technology, 1997).
- [40] D. Rothman, *Nonlinear Dynamics I: Chaos*, 2006.
- [41] R. Seydel, *From Equilibrium to Chaos: Practical Bifurcation and Stability Analysis* (Elsevier, New York, 1988).
- [42] SIAG Dynamical Systems, *DSWeb Dynamical Systems Software*, 2015.
- [43] C. Sparrow, *The Lorenz Equations: Bifurcations, Chaos and Strange Attractors* (Springer, New York, 1982).
- [44] M. C. Spruill, “Asymptotic distribution of coordinates on high dimensional spheres”, *Elect. Comm. in Probab.* **12**, 234–247 (2007).
- [45] S. H. Strogatz, *Nonlinear Dynamics and Chaos* (Westview Press, Boulder, CO, 2014).
- [46] J. M. T. Thompson and H. B. Stewart, *Nonlinear Dynamics and Chaos: Geometrical Methods for Engineers and Scientists* (Wiley, New York, 2002).
- [47] W. Tucker, “The Lorenz attractor exists”, *C. R. Acad. Sci. Paris Sér. I Math.* **328**, 1197–1202 (1999).
- [48] W. Tucker, “A rigorous ODE solver and Smale’s 14th problem”, *Found. Comp. Math.* **2**, 53–117 (2002).
- [49] M. Viana, “What’s new on Lorenz strange attractors?”, *Math. Intelligencer* **22**, 6–19 (2000).
- [50] D. Viswanath, “Symbolic dynamics and periodic orbits of the Lorenz attractor”, *Nonlinearity* **16**, 1035–1056 (2003).
- [51] D. Viswanath, “The fractal property of the Lorenz attractor”, *Physica D* **190**, 115–128 (2004).
- [52] D. Viswanath and S. Şahutoğlu, “Complex singularities and the Lorenz attractor”, *SIAM Rev.* **52**, 294–314 (2010).
- [53] E. J. Wegman and J. L. Solka, “On some mathematics for visualizing high dimensional data”, *Sankhya: Indian J. Statistics, Ser. A* **64**, 429–452 (2002).

2.6 Examples

10. Try to leave out the part that readers tend to skip.
— Elmore Leonard’s Ten Rules of Writing.

The reader is urged to study the examples collected at the end of chapters. If you want to return back to the main text, click on [click to return] pointer in the margin.

Example 2.1. A 2-dimensional vector field $v(x)$. A simple example of a flow is afforded by the unforced Duffing system

$$\begin{aligned}\dot{x}(t) &= y(t) \\ \dot{y}(t) &= -0.15 y(t) + x(t) - x(t)^3\end{aligned}\tag{2.24}$$

plotted in figure 2.4. The 2-dimensional velocity vectors $v(x) = (\dot{x}, \dot{y})$ are drawn superimposed over the configuration coordinates $(x(t), y(t))$ of state space $\mathcal{M}$, but they belong to a different space (2.10), the *tangent bundle* $\mathbf{TM}$.

Example 2.2. Lorenz strange attractor. Edward Lorenz arrived at the equation

remark 2.3

$$\dot{x} = v(x) = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} \sigma(y - x) \\ \rho x - y - xz \\ xy - bz \end{bmatrix}\tag{2.25}$$

by a drastic simplification of the Rayleigh-Benard flow. Lorenz fixed $\sigma = 10$, $b = 8/3$, and varied the “Rayleigh number” ρ . For $0 < \rho < 1$ the equilibrium $EQ_0 = (0, 0, 0)$ at the origin is attractive. At $\rho = 1$ it undergoes a pitchfork bifurcation into a pair of equilibria located at



$$x_{EQ_{1,2}} = (\pm \sqrt{b(\rho - 1)}, \pm \sqrt{b(\rho - 1)}, \rho - 1),\tag{2.26}$$

We shall not explore the Lorenz flow dependence on the ρ parameter in what follows, but here is a brief synopsis: the EQ_0 1-dimensional unstable manifold closes into a homoclinic orbit at $\rho = 13.56 \dots$. Beyond that, an infinity of associated periodic orbits are generated, until $\rho = 24.74 \dots$, where $EQ_{1,2}$ undergo a Hopf bifurcation.



All computations that follow will be performed for the Lorenz parameter choice $\sigma = 10$, $b = 8/3$, $\rho = 28$. For these parameter values the long-time dynamics is confined to the strange attractor depicted in figure 2.5, and the positions of its equilibria are marked in figure 11.2 (a). (continued in example 3.3)

return to p. 46

Example 2.3. Rössler strange attractor. The Duffing flow of figure 2.4 is bit of a bore—every orbit ends up in one of the two attractive equilibrium points. Let’s construct a flow that does not die out, but exhibits a recurrent dynamics. Start with a harmonic oscillator

$$\dot{x} = -y, \quad \dot{y} = x.\tag{2.27}$$

The solutions are re^{it} , re^{-it} , and the whole x - y plane rotates with constant angular velocity $\dot{\theta} = 1$, period $T = 2\pi$. Now make the system unstable by adding

$$\dot{x} = -y, \quad \dot{y} = x + ay, \quad a > 0,\tag{2.28}$$

or, in radial coordinates, $\dot{r} = ar \sin^2 \theta$, $\dot{\theta} = 1 + (a/2) \sin 2\theta$. The plane is still rotating with the same average angular velocity, but trajectories are now spiraling out. Any flow in the plane either escapes, falls into an attracting equilibrium point, or converges to a limit cycle. Richer dynamics requires at least one more dimension. In order to prevent the trajectory from escaping to ∞ , kick it into 3rd dimension when x reaches some value c by adding

$$\dot{z} = b + z(x - c), \quad c > 0. \quad (2.29)$$

As x crosses c , z shoots upwards exponentially, $z \simeq e^{(x-c)t}$. In order to bring it back, start decreasing x by modifying its equation to

$$\dot{x} = -y - z.$$

Large z drives the trajectory toward $x = 0$; there the exponential contraction by e^{-ct} kicks in, and the trajectory drops back toward the x - y plane. This frequently studied example of an autonomous flow is called the *Rössler flow*



$$\begin{aligned} \dot{x} &= -y - z \\ \dot{y} &= x + ay \\ \dot{z} &= b + z(x - c), \quad a = b = 0.2, \quad c = 5.7 \end{aligned} \quad (2.30)$$

(for definitiveness, we fix the parameters a, b, c in what follows). The system is as simple as they get—it would be linear, were it not for the sole bilinear term zx . Even for so ‘simple’ a system the nature of long-time solutions is far from obvious.

exercise 2.8

There are two repelling equilibrium points (2.11):

$$\begin{aligned} x_{\pm} &= \left(\frac{1}{2} \pm \frac{1}{2} \sqrt{1 - 4ab/c^2}\right)(c, -c/a, c/a) \\ x_- &\approx (ab/c, -b/c, b/c), \quad x_+ \approx (c, -c/a, c/a) \\ (x_-, y_-, z_-) &= (0.0070, -0.0351, 0.0351) \\ (x_+, y_+, z_+) &= (5.6929, -28.464, 28.464) \end{aligned} \quad (2.31)$$

One is close to the origin by construction. The other, some distance away, exists because the equilibrium condition has a 2nd-order nonlinearity.

To see what solutions look like in general, we need to resort to numerical integration. A typical numerically integrated long-time trajectory is sketched in figure 2.6 (see also figure 14.7 (a)). Trajectories that start out sufficiently close to the origin seem to converge to a *strange attractor*. We say ‘seem’ as there exists no proof that such an attractor is asymptotically aperiodic—it might well be that what we see is but a long transient on a way to an attractive periodic orbit. For now, accept that figure 2.6 and similar figures in what follows are examples of ‘strange attractors.’

(continued in exercise 2.8 and example 3.2)

(R. Paškauskas)

return to p. 46

The exercises that you should do have **underlined titles**. The rest (**smaller type**) are optional. Difficult problems are marked by any number of *** stars.

Exercises

- 2.1. **Orbits do not intersect.** An orbit in the state space $\mathcal{M}$ is the set of points one gets by evolving $x \in \mathcal{M}$ forwards and backwards in time:

$$\mathcal{M}_x = \{y \in \mathcal{M} : f^t(x) = y \text{ for } t \in \mathbb{R}\}.$$

Show that if two trajectories intersect, then they are the same curve.

- 2.2. **Evolution as a group.** The trajectory evolution f^t is a one-parameter semigroup, where (2.6)

$$f^{t+s} = f^t \circ f^s.$$

Show that it is a commutative semigroup.

In this case, the commutative character of the semigroup of evolution functions comes from the commutative character of the time parameter under addition. Can you think of any other semigroup replacing time?

- 2.3. **Almost ODE's.**

- (a) Consider the point x on $\mathbb{R}$ evolving according $\dot{x} = e^x$. Is this an ordinary differential equation?
- (b) Is $\dot{x} = x(x(t))$ an ordinary differential equation?
- (c) What about $\dot{x} = x(t+1)$?

- 2.4. **All equilibrium points are fixed points.** Show that a point of a vector field v where the velocity is zero is a fixed point of the dynamics f^t .

- 2.5. **Gradient systems.** Gradient systems (or ‘potential problems’) are a simple class of dynamical systems for which the velocity field is given by the gradient of an auxiliary function, the ‘potential’ ϕ

$$\dot{x} = -\nabla\phi(x)$$

where $x \in \mathbb{R}^d$, and ϕ is a function from that space to the reals $\mathbb{R}$.

- (a) Show that the velocity of the particle is in the direction of most rapid decrease of the function ϕ .
- (b) Show that all extrema of ϕ are fixed points of the flow.
- (c) Show that it takes an infinite amount of time for the system to reach an equilibrium point.
- (d) Show that there are no periodic orbits in gradient systems.

- 2.6. **Runge-Kutta integration.** Implement the fourth-order Runge-Kutta integration formula (see, for example, ref. [37]) for $\dot{x} = v(x)$:

$$\begin{aligned} x_{n+1} &= x_n + \frac{k_1}{6} + \frac{k_2}{3} + \frac{k_3}{3} + \frac{k_4}{6} + O(\delta\tau^5) \\ k_1 &= \delta\tau v(x_n), \quad k_2 = \delta\tau v(x_n + k_1/2) \\ k_3 &= \delta\tau v(x_n + k_2/2) \\ k_4 &= \delta\tau v(x_n + k_3). \end{aligned}$$

If you already know your Runge-Kutta, program what you believe to be a better numerical integration routine, and explain what is better about it.

- 2.7. **Rössler flow.** Use the result of exercise 2.6 or some other integration routine to integrate numerically the Rössler flow (2.30). Does the result look like a ‘strange attractor’?
- 2.8. **Equilibria of the Rössler flow.**

- (a) Find all equilibrium points (x_q, y_q, z_q) of the Rössler system (2.30). How many are there?
- (b) Assume that $b = a$. As we shall see, some surprisingly large and small numbers arise in this system. In order to understand their size, introduce parameters

$$\epsilon = a/c, \quad D = 1 - 4\epsilon^2, \quad p^\pm = (1 \pm \sqrt{D})/2.$$

Express all the equilibria in terms of $(c, \epsilon, D, p^\pm)$, expand to the first order in ϵ , and evaluate for $a = b = 0.2$, $c = 5.7$ in (2.30). In the case studied $\epsilon \approx 0.03$, these estimates are quite accurate. (continued in exercise 3.1)

(Rytis Paškauskas)

- 2.9. **Can you integrate me?** Integrating equations numerically is not for the faint of heart. It is not always possible to establish that a set of nonlinear ordinary differential equations has a solution for all times and there are many cases where the solution only exists for a limited time interval, as, for example, for the equation $\dot{x} = x^2$, $x(0) = 1$.

- (a) For what times do solutions of

$$\dot{x} = x(x(t))$$

exist? Do you need a numerical routine to answer this question?

- (b) Let's test the integrator you wrote in exercise 2.6. The equation

$$\ddot{x} = x \tag{2.32}$$

with initial conditions $x(0) = 2$ and $\dot{x} = 0$ has the solution $x(t) = e^{-t}(1 + e^{2t})$. Can your integrator reproduce this solution for the interval $t \in [0, 10]$? Check your solution by plotting the error as compared to the exact result.

- (c) Test your integrator for

$$\ddot{x} = -x \tag{2.33}$$

with the same initial conditions and integration interval.

- (d) Now we will try something a little harder. The equation is going to be third order

$$\ddot{x} + 0.6\ddot{x} + \dot{x} - |x| + 1 = 0,$$

which can be checked—numerically—to be chaotic. For initial conditions, we will always use $\ddot{x}(0) = \dot{x}(0) = x(0) = 0$. Can you reproduce the result $x(12) = 0.8462071873$ (all digits are significant)? Even though the equation being integrated is chaotic, the time intervals are not long enough for the exponential separation of trajectories to be noticeable (the exponential growth factor is ≈ 2.4).

- (e) Determine the time interval for which the solution of $\dot{x} = x^2$, $x(0) = 1$ exists.

- 2.10. **Coordinate transformations.** Changing coordinates is conceptually simple, but can become confusing when carried out in detail. The difficulty arises from confusing functional relationships, such as $x(t) = h^{-1}(y(t))$ with numerical relationships, such as $w(y) = h'(x)v(x)$. Working through an example will clear this up.

- (a) The differential equation in $\mathcal{M}$ is $\dot{x} = \{2x_1, x_2\}$ and the change of coordinates from $\mathcal{M}$ to $\mathcal{M}'$ is $h(x_1, x_2) = \{2x_1 + x_2, x_1 - x_2\}$. Solve for $x(t)$. Find h^{-1} .
- (b) Show that in the transformed space $\mathcal{M}'$, the differential equation is

$$\frac{d}{dt} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 5y_1 + 2y_2 \\ y_1 + 4y_2 \end{bmatrix}.$$

Solve this system. Does it match the solution in the $\mathcal{M}$ space?

- 2.11. **Deterministic collinear helium dynamics.** In order to apply periodic orbit theory to the quantization of helium, we need to compute deterministic periodic orbits of the helium system. We begin by evaluating periodic orbits for the colinear helium atom (8.32)

$$H = \frac{1}{2}p_1^2 + \frac{1}{2}p_2^2 - \frac{Z}{r_1} - \frac{Z}{r_2} + \frac{1}{r_1 + r_2}.$$

The nuclear charge for helium is $Z = 2$. Colinear helium has only 3 degrees of freedom, and the dynamics can be visualized as motion in the (r_1, r_2) , $r_i \geq 0$ quadrant. In (r_1, r_2) -coordinates the potential is singular for $r_i \rightarrow 0$ nucleus-electron collisions. These 2-body collisions can be regularized by rescaling the coordinates, with details given in sect. A2.2. In the transformed coordinates (x_1, x_2, p_1, p_2) the Hamiltonian equations of motion take the form

$$\begin{aligned} \dot{P}_1 &= 2Q_1 \left[2 - \frac{P_2^2}{8} - Q_2^2 \left(1 + \frac{Q_2^2}{R^4} \right) \right] \\ \dot{P}_2 &= 2Q_2 \left[2 - \frac{P_1^2}{8} - Q_1^2 \left(1 + \frac{Q_1^2}{R^4} \right) \right] \\ \dot{Q}_1 &= \frac{1}{4}P_1Q_2^2, \quad \dot{Q}_2 = \frac{1}{4}P_2Q_1^2. \end{aligned} \quad (2.34)$$

where $R = (Q_1^2 + Q_2^2)^{1/2}$.

Integrate the equations of motion by the fourth order Runge-Kutta computer routine of exercise 2.6 (or whatever integration routine you like). A convenient way to visualize the 3-dimensional state space orbit is by projecting it onto the 2-dimensional $(r_1(t), r_2(t))$ plane. (continued in exercise 3.4) (Gregor Tanner, Per Rosenqvist)

- 2.12. **Surface area of a unit sphere.** Compute the volume of a unit sphere in d dimensions.
- 2.13. **In high dimensions any two vectors are (nearly) orthogonal.** Among humble plumbers laboring with extremely high-dimensional ODE discretizations of fluid and other PDEs, there is an inclination to visualize the ∞ -dimensional state space flow by projecting it onto a basis constructed from a few random coordinates, let's say the 2nd Fourier mode along the spatial x direction against the 4th Chebyshev mode along the y direction. It's easy, as these are typically the computational degrees of freedom. As we will now show, it's easy but not smart, with vectors representing the dynamical states of interest being almost orthogonal to any such random basis.

Suppose your state space $\mathcal{M}$ is a real 10 247-dimensional vector space, and you pick from it two vectors $x_1, x_2 \in \mathcal{M}$ at random. What is the angle between them likely to be?

By asking for 'angle between two vectors' we have implicitly assumed that there exist is a dot product

$$x_1^\top \cdot x_2 = \|x_1\| \|x_2\| \cos(\theta_{12}),$$

so let's make these vectors unit vectors, $\|x_j\| = 1$. When you think about it, you would be hard put to say what 'uniform probability' would mean for a vector $x \in \mathcal{M} = \mathbb{R}^{10247}$, but for a unit vector it is obvious: probability that x direction lies within a solid angle $d\Omega$ is $d\Omega/(\text{unit hyper-sphere surface})$.

So what is the surface of the unit sphere (or, the total solid angle) in d dimensions? One way to compute it is to evaluate the Gaussian integral

$$I_d = \int_{-\infty}^{\infty} dx_1 \cdots dx_d e^{-\frac{1}{2}(x_1^2 + \cdots + x_d^2)} \quad (2.35)$$

in cartesian and polar coordinates. Show that

- (a) In cartesian coordinates $I_d = (2\pi)^{d/2}$.
- (b) Recast the integrals in polar coordinate form. You know how to compute this integral in 2 and 3 dimensions. Show by induction that the surface S_{d-1} of unit d -ball, or the total solid angle in even and odd dimensions is given by

$$S_{2k} = \frac{2(2\pi)^k}{(2k-1)!!}, \quad S_{2k+1} = \frac{2\pi^{k+1}}{k!}. \quad (2.36)$$

- (c) Show, by examining the form of the integrand in the polar coordinates, that for arbitrary, perhaps even complex dimension $d \in \mathbb{C}$

$$S_{d-1} = 2\pi^{d/2}/\Gamma(d/2).$$

(In Quantum Field Theory integrals over 4-momenta are brought to polar form and evaluated as functions of a complex dimension parameter d . This procedure is called the ‘dimensional regularization’.)

- (d) Check your formula for $d = 2$ (1-sphere, or the circle) and $d = 3$ (2-sphere, or the sphere).
- (e) Determine the limit of S_d for large d . (Hint: it’s not what you think. Try Sterling’s formula).

So now that we know the volume of a sphere, what is the most likely angle between two vectors x_1, x_2 picked at random? We can rotate coordinates so that x_1 is aligned with the ‘z-axis’ of the hypersphere. An angle θ then defines a meridian around the ‘z-axis’.

- (f) Show that probability $P(\theta)d\theta$ of finding two vectors at angle θ is given by the area of the meridional strip of width $d\theta$, and derive the formula for it:

$$P(\theta) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(d/2)}{\Gamma((d-1)/2)}.$$

(One can write an analytic expression for this in terms of Beta functions, but it is unnecessary for the problem at hand.)

- (g) Show that for large d the probability $P(\theta)$ tends to a normal distribution with mean $\theta = \pi/2$ and variance $1/d$.

So, in d -dimensional vector space the two random vectors are nearly orthogonal, within accuracy of $\theta = \pi/2 \pm 1/d$.

If you want to learn more, the [lecture notes](#) by Hermann Flaschka, entitled *Some geometry in high-dimensional spaces*, offer a high quality solution to this exercise.

If you are a humble plumber, and the notion of a vector space is some abstract hocus-pocus to you, try thinking this way. Your 2nd Fourier mode basis vector is something that wiggles twice along your computation domain. Your turbulent state is very wiggly. The product of the two functions integrated over the computational domain will average to zero, with a small leftover. We have just estimated that with dumb choices of coordinate bases this leftover will be of order of $1/10\,247$, which is embarrassingly small for displaying a phenomenon of order ≈ 1 .

Several intelligent choices of coordinates for state space projections are described in Gibson *et al.* [21] and the web tutorial ChaosBook.org/tutorials.

Sara A. Solla and P. Cvitanović

Chapter 3

Discrete time dynamics

Gentles, perchance you wonder at this show; But wonder on, till truth make all things plain.

— W. Shakespeare, *A Midsummer Night's Dream*

THE TIME PARAMETER in the definition of a dynamical system can be either continuous or discrete. Discrete time dynamical systems arise naturally from flows. In general there are two strategies for replacing a continuous-time flow by iterated mappings; one can cut it into Poincaré sections or use *strobing* to capture a sequence of instants in time. Think of your partner moving to the beat in a disco: a sequence of frozen stills. While ‘strobing’ is what any numerical integrator does, by representing a trajectory by a sequence of time-integration step separated points, strobing is in general not a reduction of a flow, as the sequence of strobed points still resides in the full state space $\mathcal{M}$, of dimensionality d . An exception are non-autonomous flows that are externally periodically forced. In that case it might be natural to observe the flow by strobing it at time intervals fixed by an external forcing, as in example 8.7 where strobing of a periodically forced Hamiltonian leads to the ‘standard map.’

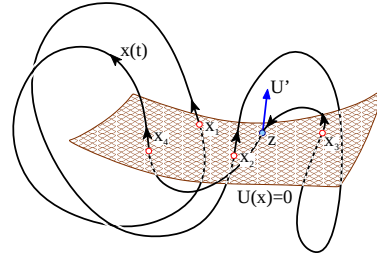
section 2.1

example 8.7

In the *Poincaré section method*, one records the coordinates of a trajectory whenever the trajectory crosses a prescribed trigger. This triggering event can be as simple as the vanishing of one of the coordinates, or as complicated as the trajectory cutting through a curved hypersurface. A Poincaré section (or, for the remainder of ChaosBook, often just ‘section’) is *not* a projection onto a lower-dimensional space: rather, it is a local change of coordinates to a direction along the flow, and the remaining coordinates (spanning the section) transverse to it. No information about the flow is lost by reducing it to a set of Poincaré section points and the return maps connecting them. The full space trajectory can always be reconstructed by integration from the nearest point in the section.

Reduction of a continuous time flow to its Poincaré section is a powerful visualization tool. But, the method of sections is more than a visualization tool; it is also a fundamental tool of dynamics. To fully unravel the geometry of a

Figure 3.1: A trajectory $x(t)$ that intersects a Poincaré section $\mathcal{P}$ at times t_1, t_2, t_3, t_4 , and closes a cycle $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4)$, $\hat{x}_k = x(t_k) \in \mathcal{P}$ of topological period 4 with respect to the section. In general, the intersections are not normal to the section. Note also that the crossing z does not count, as it is in the wrong direction.



chaotic flow, one *has to* quotient all of its symmetries, and time evolution is just one of these symmetries. (This delphic piece of hindsight will be illuminated in chapter 12.)

3.1 Poincaré sections



A continuous time flow decomposes the state space into Lagrangian ‘spaghetti’ of figure 2.2, a union of non-intersecting 1-dimensional orbits. Any point (2.3) on an orbit can be used to label the orbit. Thus, the state space is reduced to a ‘skew-product’ of a $(d-1)$ -dimensional space $\mathcal{P}$ of labeling points $\hat{x}_j \in \mathcal{P}$ and the corresponding 1-dimensional orbit curves $\mathcal{M}_j$ on which the flow acts as a time translation. However, orbits can be arbitrarily complicated and, if unstable, uncontrollable for times beyond the Lyapunov time (1.1). In practice, it is necessary to split the orbit into finite trajectory segments, with time intervals corresponding to the shortest recurrence times on a non-wondering set of the flow, i.e. *finite* times for which the flow is computable.



A particular prescription for picking the orbit-labeling points is called a *Poincaré section*. In introductory texts Poincaré sections are treated as pretty visualizations of chaotic flow, but their dynamical significance is much deeper than that. Once a section is defined, a ‘Lagrangian’ description of the flow (discussed above, page 47) is replaced by the ‘Eulerian’ formulation, with the trajectory-tangent velocity field $v(\hat{x})$, $\hat{x} \in \mathcal{P}$, enabling us to go freely between the time-quotiented space $\mathcal{P}$ and the full state space $\mathcal{M}$. The *transverse dynamics* of this system is important and describes how nearby trajectories attract/repeal each other. Also, transverse dynamics is used to map P of $\mathcal{P} \rightarrow \mathcal{P}$ induced by the flow. In contrast, dynamics *along* orbits is of secondary importance.

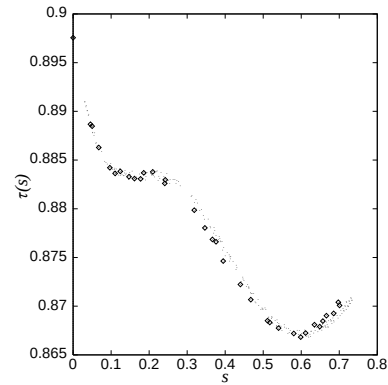
chapter 12

Successive trajectory intersections with a Poincaré section, a $(d-1)$ -dimensional hypersurface embedded in the d -dimensional state space $\mathcal{M}$, figure 3.1, define the *return map*, or in the rest of ChaosBook, simply return map $P(\hat{x})$, a $(d-1)$ -dimensional map of form

$$\hat{x}' = P(\hat{x}) = f^{\tau(\hat{x})}(\hat{x}), \quad \hat{x}', \hat{x} \in \mathcal{P}. \quad (3.1)$$

Here the *first return function* $\tau(\hat{x})$ —sometimes referred to as the *ceiling function*—is the time of flight to the next section for a trajectory starting at $\hat{x}$, see figure 3.2. The choice of the section hypersurface $\mathcal{P}$ is altogether arbitrary. It is rarely possible to define a single section that cuts across all trajectories of interest. Fortunately,

Figure 3.2: The return time $\tau(s)$ as a function of the parameter s for the Kuramoto-Sivashinsky system discussed in chapter 30, evaluated on the periodic points, as in figure 30.10, with the diamonds obtained by 34 periodic points and the dots by 240 periodic points. The fine structure is due to the fractal structure of the attractor (from ref. [2]).



one often needs only a local section in the neighborhood of a *template* point, a finite hypersurface of codimension 1 intersected by a swarm of trajectories near to the trajectory of interest (the case of several sections is discussed in sect. 15.6). Such a hypersurface can be specified implicitly by a single condition, through a function $U(x)$ that is zero whenever a point x is on the Poincaré section,

$$\hat{x} \in \mathcal{P} \quad \text{iff} \quad U(\hat{x}) = 0. \quad (3.2)$$

The gradient of $U(x)$ evaluated at $\hat{x} \in \mathcal{P}$ serves a two-fold function. First, the flow should pierce the hypersurface $\mathcal{P}$, rather than being tangent to it. A nearby point $\hat{x} + \delta x$ is in the hypersurface $\mathcal{P}$ if $U(\hat{x} + \delta x) = 0$. A nearby point on the trajectory is given by $\delta x = v\delta t$, so a traversal is ensured by the *transversality condition*

$$(v \cdot \nabla U) = \sum_{j=1}^d v_j(\hat{x}) \partial_j U(\hat{x}) \neq 0, \quad \partial_j U(\hat{x}) = \frac{\partial}{\partial \hat{x}_j} U(\hat{x}), \quad \hat{x} \in \mathcal{P}. \quad (3.3)$$

Second, the gradient ∇U defines the orientation of the hypersurface $\mathcal{P}$. The flow is oriented as well, and a periodic orbit can pierce $\mathcal{P}$ twice, traversing it in either direction, as in figure 3.1. Hence the definition of return map $P(\hat{x})$ needs to be supplemented with the orientation condition

$$\begin{aligned} \hat{x}_{n+1} &= P(\hat{x}_n), & U(\hat{x}_{n+1}) &= U(\hat{x}_n) = 0, & n &\in \mathbb{Z}^+ \\ & & \sum_{j=1}^d v_j(\hat{x}_n) \partial_j U(\hat{x}_n) &> 0. \end{aligned} \quad (3.4)$$

In this way the continuous time t flow $x(t) = f^t(x)$ is reduced to a discrete time n sequence $\hat{x}_n$ of successive *oriented* trajectory traversals of $\mathcal{P}$.

The simplest choice of a Poincaré section is a hyperplane $\mathcal{P}$ specified by a template point (an important state of the system, located at the tip of the vector $\hat{x}'$) and a normal vector $\hat{n}$ perpendicular to the hyperplane. A Poincaré section point $\hat{x}$ is in the hyperplane, if it satisfies the linear condition

$$\hat{x} \in \mathcal{P} \quad \text{iff} \quad U(\hat{x}) = (\hat{x} - \hat{x}') \cdot \hat{n} = 0. \quad (3.5)$$

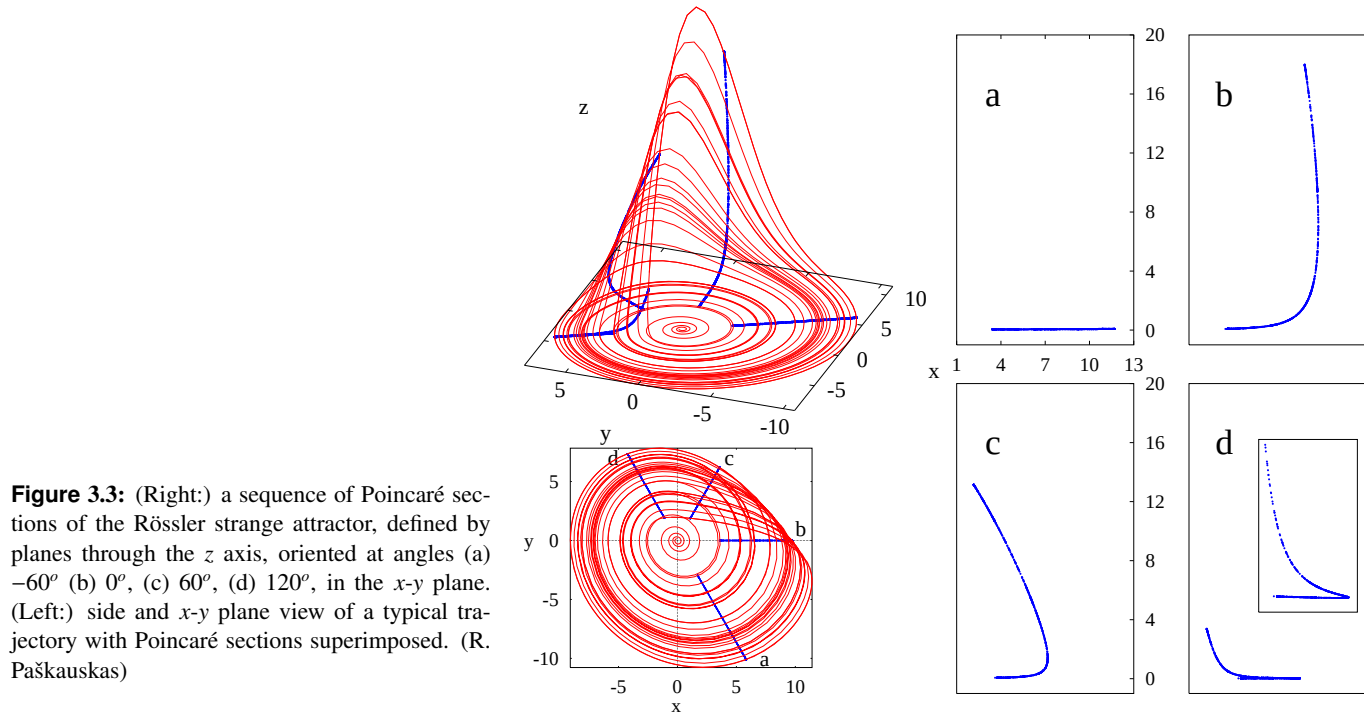


Figure 3.3: (Right:) a sequence of Poincaré sections of the Rössler strange attractor, defined by planes through the z axis, oriented at angles (a) -60° (b) 0° , (c) 60° , (d) 120° , in the x - y plane. (Left:) side and x - y plane view of a typical trajectory with Poincaré sections superimposed. (R. Paškauskas)

Consider a circular periodic orbit centered at $\hat{x}'$, but not lying in $\mathcal{P}$. It pierces the hyperplane twice; the $v \cdot \hat{n} > 0$ traversal orientation condition (3.4) ensures that the first return time is the full period of the cycle. The simplest choice of the hyperplane orientation is to chose it to be normal to the tangent of the trajectory passing through the template point, i.e., the state space velocity,

example 15.2

$$\hat{x} \in \mathcal{P} \quad \text{iff} \quad U(\hat{x}) = (\hat{x} - \hat{x}') \cdot v(\hat{x}') = 0. \quad (3.6)$$

With a sufficiently clever choice of a Poincaré section or a set of sections, any orbit of interest intersects a section, see figure 3.3. Depending on the application, one might need to convert the discrete time n back to the continuous flow time. This is accomplished by adding up the first return function times $\tau(\hat{x}_n)$, with the accumulated flight time given by

$$t_{n+1} = t_n + \tau(\hat{x}_n), \quad t_0 = 0, \quad x_n \in \mathcal{P}. \quad (3.7)$$

Other quantities integrated along the trajectory can be defined in a similar manner, and will need to be evaluated in the process of evaluating dynamical averages.

chapter 20

A few examples may help visualize this. A typical trajectory of the 3-dimensional Rössler flow is plotted in figure 2.6. A sequence of Poincaré sections of figure 3.3 illustrates the ‘stretch & fold’ action of Rössler flow. Figure 3.4 exhibits a set of return maps (3.1).



example 3.1
p. 81



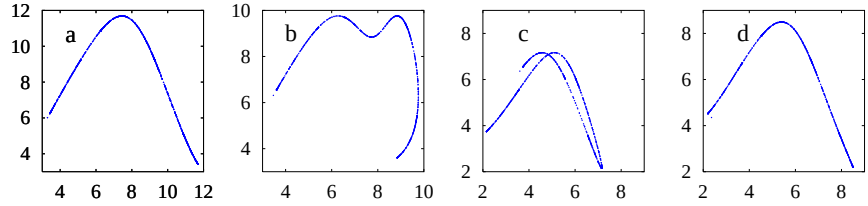
example 3.2
p. 81



The above examples illustrate why a Poincaré section gives a more informative snapshot of the flow than the full flow portrait. For example, while the full flow



Figure 3.4: Return maps for the $r_n \rightarrow r_{n+1}$ radial distance Poincaré sections of figure 3.3. The ‘multi-valuedness’ of (b) and (c) is only apparent: the full return map is 2-dimensional, $\{r', z'\} = P\{r, z\}$. (R. Paškauskas)



portrait of the Rössler flow figure 2.6 gives us no sense of the thickness of the attractor, we see clearly in the Poincaré sections of figure 3.3 that even though the return maps are 2-dimensional $\rightarrow$ 2-dimensional, the flow contraction is so strong that for all practical purposes it renders the return maps 1-dimensional. (We shall quantify this claim in example 4.5.)



question 3.1



fast track:
sect. 3.3, p. 75

3.1.1 Section border



How far does the neighborhood of a template $\hat{x}'$ extend along the hyperplane (3.5)? A section captures faithfully neighboring orbits as long as it cuts them transversally; it fails the moment the velocity field at point $\hat{x}^*$ fails to pierce the section. At this location the velocity is tangent to the section and, thus, orthogonal to the template normal $\hat{n}$,



$$\hat{n} \cdot v(\hat{x}^*) = 0, \quad \hat{x}^* \in S, \quad (3.8)$$

which shows that the normal component of $v(\hat{x})$ vanishes at $\hat{x}^*$. For a smooth flow, such points form a smooth $(d-2)$ -dimensional *section border* $S \subset \mathcal{P}$, encompassing the open neighborhood of the template characterized by qualitatively similar flow. We shall refer to this region of the section hyperplane as the (maximal) chart of the template neighborhood for a given hyperplane (3.5).

If the template point is an equilibrium x_q , no dynamics exist and velocity vanishes ($v(x_q) = 0$ by definition of equilibrium). In this case velocity cannot be used to define a normal to the section, but instead we use the local linearized flow to construct the local Poincaré section $\mathcal{P}$. We orient $\mathcal{P}$ so that unstable eigenvectors are transverse to the section, and the slowest contracting eigenvector is tangent to the section, as in figure 4.6. This ensures that the flow is transverse to $\mathcal{P}$ in an open neighborhood of the template x_q .

exercise 3.7

Visualize the flow as a smooth 3-dimensional steady fluid flow cut by a 2-dimensional sheet of light. Lagrangian particle trajectories either cross, are tangent to, or fail to reach this plane; the 1-dimensional curves of tangency points define the section border. An example is offered by the velocity field of the Rössler flow of figure 4.5. Pick a Poincaré section hyperplane so it goes through both equilibrium points. The section might be transverse to a large neighborhood around the inner equilibrium x_- , but dynamics around the outer equilibrium x_+ is totally

different. The competition between these two types of motion is likely to lead to the vanishing of $v_{\perp}(\hat{x})$, the component of $v(\hat{x})$ normal to the section, someplace in-between the two equilibria. A section is good up to the section border, but beyond it an orbit infinitesimally close to $\hat{x}^*$ generically does not cross the section hyperplane.

For 3-dimensional flows, the section border $\mathcal{S}$ is a 1-dimensional closed curve in the section 2-dimensional $\mathcal{P}$, and easy to visualize. In higher dimensions, the section border is a $(d-2)$ -dimensional manifold, not easily visualized, and the best one can do is to keep checking for change of sign (3.4) at Poincaré section returns of nearby trajectories close to the section border hypersurface $\mathcal{S}$; (3.8) will be positive inside, negative immediately outside $\mathcal{S}$.

Thus for a nonlinear flow, with its complicated curvilinear invariant manifolds, a single section rarely suffices to capture all of the dynamics of interest.

3.1.2 What is the best Poincaré section?



In practice, picking sections is a dark and painful art, especially for high-dimensional flows where the human visual cortex falls short. It helps to understand why we need them in the first place.

Whenever a system has a continuous symmetry G , any two solutions related by the symmetry are equivalent. We do not want to keep recomputing these over and over. We would rather replace the whole continuous family of solutions by one solution in order to be more efficient. This approach replaces the dynamics $(\mathcal{M}, f)$ with dynamics on the *quotient state space* $(\mathcal{M}/G, \hat{f})$. For now, we only remark that constructing explicit quotient state space flow $\hat{f}$ is either extremely difficult, impossible, or generates unintelligible literature. Our solution (see chapter 12) will be to resort to the method of slices.

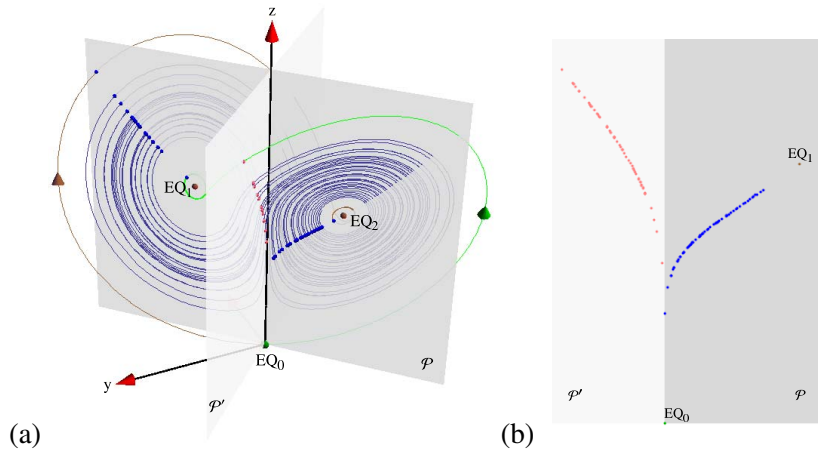
chapter 12

Time evolution itself is a 1-parameter Lie group, albeit a highly nontrivial one (otherwise ChaosBook would not be as much of a doorstop). The invariants of the flow are its infinite-time orbits; particularly useful invariants are compact orbits such as equilibrium points, periodic orbits, and tori. For any orbit it suffices to pick a single state space point $x \in \mathcal{M}_p$. The rest of the orbit is generated by the flow.

Choice of this one ‘labeling’ point is utterly arbitrary; in dynamics this is called a ‘Poincaré section’, and in theoretical physics this goes by the exceptionally uninformative name of ‘gauge fixing’. The price is that one generates ‘ghosts’, or, in dynamics, the price is that increasing dimensionality of the state space adds further constraints (see sect. 7.2). Gauge fixing is a commonly deployed but inelegant procedure where symmetry is broken for computational convenience, and restored only at the end of the calculation, when all broken pieces are reassembled.

With this said, there are a few rules of thumb to follow: (a) You can pick as

Figure 3.5: (a) Lorenz flow figure 2.5 cut by $y = x$ Poincaré section plane $\mathcal{P}$ through the z axis and both $EQ_{1,2}$ equilibria. Points where flow pierces into section are marked by dots. To aid visualization of the flow near the EQ_0 equilibrium, the flow is cut by the second Poincaré section, $\mathcal{P}'$, through $y = -x$ and the z axis. (b) Poincaré sections $\mathcal{P}$ and $\mathcal{P}'$ laid side-by-side. The singular nature of these sections close to EQ_0 will be elucidated in example 4.6 and figure 14.14 (b). (E. Siminos)



many sections as convenient, as discussed in sect. 15.6. (b) For ease of computation, pick linear sections (3.5) when possible. (c) If equilibria play important role in organizing a flow, pick sections that go through them (see example 3.3). In that case, try to place contracting eigenvectors inside the hyperplane, see Lorenz figure 3.5. Remember, the stability eigenvectors are never orthogonal to each other, unless that is imposed by some symmetry. (d) If you have a global discrete or continuous symmetry, pick sections left invariant by the symmetry (see example 11.8). For example, setting the normal vector $\hat{n}$ in (3.5) at x to be the velocity $v(x)$ is natural and locally transverse. (e) If you are solving a local problem, like finding a periodic orbit, you do not need a global section. Pick a section or a set of (multi-shooting) sections on the fly, requiring only that they are locally transverse to the flow. (f) If you have another rule of thumb dear to you, let us know.



chapter 11



example 3.3
p. 81

3.2 Computing a Poincaré section

(R. Mainieri)



For almost any flow of physical interest, a Poincaré section is not available in analytic form, so one tends to determine it crudely, by numerically bracketing the trajectory traversals of a section and iteratively narrowing the bracketing time interval. We describe here a smarter method, which you will only need when you seriously look at a strange attractor, with millions of points embedded in a high(er)-dimensional Poincaré section - so skip this section on the first reading.

remark 3.2

Consider the system (2.9) of ordinary differential equations in the vector variable $x = (x_1, x_2, \dots, x_d)$

$$\frac{dx_i}{dt} = v_i(x, t), \quad (3.9)$$

where the flow velocity v is a vector function of the position in state space x and time t . In general, the map $f^{\tau_n}(x_n) = x_n + \int d\tau v(x(\tau))$ cannot be integrated ana-

lytically, so we resort to numerical integration to determine the trajectories of the system. Our task is to determine the points at which the numerically integrated trajectory traverses a given hypersurface. The hypersurface will be specified implicitly through a function $U(x)$ that is zero, whenever a point x is on the Poincaré section, such as the hyperplane (3.5).

If we use a tiny step size in our numerical integrator, we can observe the value of U as we integrate; its sign will change as the trajectory crosses the hypersurface. The problem with this method is that we have to use a very small integration time step. However, there is a better way to land exactly on the Poincaré section.

Let t_a be the time just before U changes sign, and t_b the time just after it changes sign. The method for landing exactly on the Poincaré section will be to convert one of the space coordinates into an integration variable for the part of the trajectory between t_a and t_b . Using

$$\frac{dx_k}{dx_1} \frac{dx_1}{dt} = \frac{dx_k}{dx_1} v_1(x, t) = v_k(x, t), \quad (3.10)$$

we can rewrite the equations of motion (3.9) as

$$\frac{dt}{dx_1} = \frac{1}{v_1}, \dots, \frac{dx_d}{dx_1} = \frac{v_d}{v_1}. \quad (3.11)$$

Now we use x_1 as the ‘time’ in the integration routine and integrate it from $x_1(t_a)$ to the value of x_1 on the hypersurface, determined by the hypersurface intersection condition (3.5). This is the end point of the integration, with no need for any interpolation or backtracking to the surface of the section. The x_1 -axis need not be perpendicular to the Poincaré section; any x_i can be chosen as the integration variable, provided the x_i -axis is not parallel to the Poincaré section at the trajectory intersection point. If the section crossing is transverse (3.3), v_1 cannot vanish in the short segment bracketed by the integration step preceding the section, and the point on the Poincaré section.



example 3.4
p. 82

3.3 Mappings

Do it again! (and again! and again! and ...)
—Isabelle, age 3



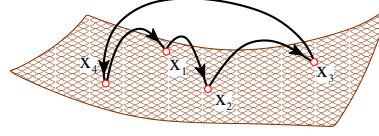
We have motivated discrete time dynamics by considering sections of a continuous flow and reduced the continuous-time flow to a family of maps $P(\hat{x})$, mapping points $\hat{x}$ from a section to a section. Many settings exist in which dynamics is inherently discrete, and naturally described by repeated iterations of the same map

$$f : \mathcal{M} \rightarrow \mathcal{M},$$



remark 3.1

Figure 3.6: A flow $x(t)$ of figure 3.1 represented by a return map that maps points in the Poincaré section $\mathcal{P}$ as $\hat{x}_{n+1} = f(\hat{x}_n)$. In this example the orbit of $\hat{x}_1$ is periodic and consists of the four periodic points $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4)$.



or sequences of consecutive applications of a finite set of maps, a different map, $f_A, f_B, \dots$, for points in different regions $\{\mathcal{M}_A, \mathcal{M}_B, \dots, \mathcal{M}_Z\}$,

$$\{f_A, f_B, \dots, f_Z\} : \mathcal{M} \rightarrow \mathcal{M}. \quad (3.12)$$

For example, such maps relate different sections among a set of Poincaré sections. The discrete ‘time’ is then an integer, the number of applications of the map or maps. Since explicitly writing out formulas involving repeated applications of a set of maps can be awkward, we streamline the notation by denoting the (non-commutative) map composition by ‘ $\circ$ ’

$$f_Z(\dots f_B(f_A(x)) \dots) = f_Z \circ \dots \circ f_B \circ f_A(x), \quad (3.13)$$

and the n th iterate of map f by

$$f^n(x) = f \circ f^{n-1}(x) = f(f^{n-1}(x)), \quad f^0(x) = x.$$

The *trajectory* of x is the finite set of points

$$\{x, f(x), f^2(x), \dots, f^n(x)\},$$

section 2.1

traversed in time n , and $\mathcal{M}_x$, the *orbit* of x , is the subset of all points of $\mathcal{M}$ that can be reached by iterations of f . A *periodic point* (cycle point) x_k belonging to a *periodic orbit* (cycle) of period n is a real solution of

$$f^n(x_k) = f(f(\dots f(x_k) \dots)) = x_k, \quad k = 0, 1, 2, \dots, n-1. \quad (3.14)$$

For example, the orbit of $\hat{x}_1$ in figure 3.6 is a set of four cycle points, $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4)$.

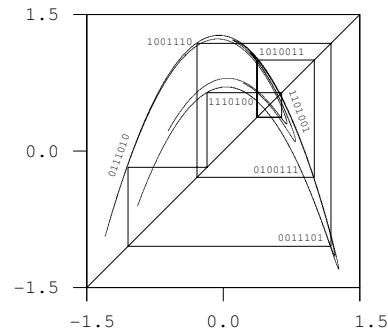
The functional form of such return maps P as figure 3.4 can be approximated by tabulating the results of integration of the flow from $\hat{x}$ to the first Poincaré section return for many $\hat{x} \in \mathcal{P}$, and constructing a function that interpolates through these points. If we find a good approximation to $P(\hat{x})$, we can get rid of numerical integration altogether, by replacing the continuous time trajectory $f^t(\hat{x})$ by iteration of the return map $P(\hat{x})$. Constructing accurate $P(\hat{x})$ for a given flow can be tricky, but we can already learn much from approximate return maps. Multinomial approximations

$$P_k(\hat{x}) = a_k + \sum_{j=1}^d b_{kj} \hat{x}_j + \sum_{i,j=1}^d c_{kij} \hat{x}_i \hat{x}_j + \dots, \quad \hat{x} \in \mathcal{P} \quad (3.15)$$

to return maps

$$\begin{bmatrix} \hat{x}_{1,n+1} \\ \hat{x}_{2,n+1} \\ \vdots \\ \hat{x}_{d,n+1} \end{bmatrix} = \begin{bmatrix} P_1(\hat{x}_n) \\ P_2(\hat{x}_n) \\ \vdots \\ P_d(\hat{x}_n) \end{bmatrix}, \quad \hat{x}_n, \hat{x}_{n+1} \in \mathcal{P}$$

Figure 3.7: The strange attractor and an unstable period 7 cycle of the Hénon map (3.18) with $a = 1.4$, $b = 0.3$. The periodic points in the cycle are connected to guide the eye. (from K.T. Hansen [6])



motivate the study of model mappings of the plane, such as the Hénon map and the Lozi map.



example 3.5
p. 82



example 3.6
p. 83

What we get by iterating such maps is—at least qualitatively—not unlike what we get from Poincaré section of flows such as the Rössler flow figure 3.4. For an arbitrary initial point this process might converge to a stable limit cycle, to a strange attractor, to a false attractor (due to roundoff errors), or diverge. In other words, mindless iteration is essentially uncontrollable, and we will need to resort to more thoughtful explorations. As we shall explain in due course, strategies for systematic exploration rely on stable/unstable manifolds, periodic points, saddle-straddle methods and so on.



exercise 6.3



example 3.7
p. 83

As we shall see in sect. 14.3, an understanding of 1-dimensional dynamics is indeed the essential prerequisite to unraveling the qualitative dynamics of many higher-dimensional dynamical systems. For this reason many expositions of the theory of dynamical systems commence with a study of 1-dimensional maps. We prefer to stick to flows, as that is where the physics is.

appendix A10.3



fast track:
sect. 4, p. 86

Résumé

In recurrent dynamics a trajectory exits a region in state space and then reenters it infinitely often, with finite return times. If the orbit is periodic, it returns after a full period. So, on average, nothing much really happens along the trajectory, but the behavior of neighboring trajectories transverse to the flow is quite important. A visualization of a strange attractor can be greatly facilitated by a felicitous choice of Poincaré sections, and the computation greatly sped up by a reduction of flow to return maps. This observation motivates in turn the study of discrete-time dynamical systems generated by iterations of maps.

A particularly natural application of the Poincaré section method is the reduction of a billiard flow to a boundary-to-boundary return map, described in chapter 9. As we show in appendix A2, further simplification of a return map, or any nonlinear map, can be attained through rectifying these maps locally by means of smooth conjugacies.

chapter 9
appendix A2

In truth, as we shall see in chapter 12, the reduction of continuous time flow by the method of Poincaré sections is not a convenience, but an absolute necessity. In order to make sense of an ergodic flow, all of its continuous symmetries must be reduced, including the evolution in time.

Commentary

Remark 3.1. Functions, maps, mappings. In mathematics, ‘mapping’ is a noun, ‘map’ is a verb. Nevertheless, ‘mapping’ is often shortened to ‘map’ and is often used as a synonym for ‘function.’ ‘Function’ is used for mappings that map to a single point in $\mathbb{R}$ or $\mathbb{C}$, while a mapping which maps to $\mathbb{R}^d$ would be called a ‘mapping,’ and not a ‘function.’ Likewise, if a point maps to several points and/or has several pre-images, this is a ‘many-to-many’ mapping, rather than a function. In his review [14], Smale refers to iterated maps as ‘diffeomorphisms’, in contradistinction to ‘flows’, which are 1-parameter groups of diffeomorphisms. In this sense, the dynamical evolution of an initial state to a state finite time later is a (time-forward) map.

Remark 3.2. Determining a Poincaré section. The trick described in sect. 3.2 is accredited to Hénon [7, 10, 15]. The idea of changing the integration variable from time to one of the coordinates, although simple, avoids the alternative of having to interpolate the numerical solution to determine the intersection.

Question 3.1. Alice wants to know

Q Why does a Poincaré section have to be a hypersurface of codimension 1?

A In 2 dimensions a curve can intersect a 1-dimensional Poincaré section curve in a point, but it has zero probability of intersecting a random 0-dimensional point in the plane. In 3 dimensions a line intersects a 2-dimensional plane in a single point, but it has zero probability of intersecting a random 1-dimensional line, or a random 0-dimensional point. In 4 dimensions a line intersects a 3-dimensional hyperplane (a volume) in a single point, but it has zero probability of intersecting a random 2-dimensional plane. You need a Poincaré section to separate, at least locally, the d -dimensional state space above the section hypersurface from the state space below, and only a codimension $1 = (d - 1)$ -dimensional hypersurface can do that. If a trajectory were a 2-dimensional ribbon or a tube, it would intersect a Poincaré section of codimension 2 in a point, but a generic codimension 3 hypersurface would not cut it at all.

Remark 3.3. Hénon, Lozi maps. The Hénon map is of no particular physical import in and of itself—its significance lies in the fact that it is a minimal normal form for modeling flows near a saddle-node bifurcation, and that it is a prototype of the stretching and folding dynamics that leads to deterministic chaos. It is generic in the sense that it can exhibit arbitrarily complicated symbolic dynamics and mixtures of hyperbolic and non-hyperbolic behaviors. Its construction was motivated by the best known early example of

‘deterministic chaos,’ the Lorenz equation, see example 2.2 and remark 2.3. Y. Pomeau’s studied the Lorenz attractor on an analog computer, and his insights into its stretching and folding dynamics motivated Hénon [9] to introduce the Hénon map in 1976. Hénon’s and Lorenz’s original papers can be found in reprint collections refs. [3, 8]. They are a pleasure to read, and these articles are still the best introduction to the physics motivating such models. Hénon [9] conjectured that a generic initial point converges to a *strange attractor*, for an $(a, b) = (1.4, 0.3)$ Hénon map. Its existence has never been proven. While for all practical purposes this is a strange attractor, it has not been demonstrated that long time iterations are not attracted by some long attracting limit cycle. Indeed, the pruning front techniques that we describe below enable us to find stable attractors arbitrarily close by in the parameter space, such as the 13-cycle attractor at $(a, b) = (1.39945219, 0.3)$. A rigorous proof of the existence of Hénon attractors close to 1-dimensional parabola map was developed by Benedicks and Carleson [1]. A detailed description of the dynamics of the Hénon map is given by Mira and coworkers [4, 5, 12], as well as many other authors. The Lozi map (3.20) is particularly convenient in investigating the symbolic dynamics of 2-dimensional mappings. Both the Lorenz and Lozi [11] systems uniformly expand smooth systems with singularities. The existence of the attractor for the Lozi map was proven by M. Misiurewicz [13], and the existence of the SRB measure was established by L.-S. Young [16].

exercise 6.3

section 19.1

References

- [1] M. Benedicks and L. Carleson, “On iterations of $1 - ax^2$ on $(-1, 1)$ ”, *Ann. Math.* **122**, 1–25 (1985).
- [2] F. Christiansen, P. Cvitanović, and V. Putkaradze, “Hopf’s last hope: Spatiotemporal chaos in terms of unstable recurrent patterns”, *Nonlinearity* **10**, 55–70 (1997).
- [3] P. Cvitanović, *Universality in Chaos*, 2nd ed. (Adam Hilger, Bristol, 1989).
- [4] D. Fournier, H. Kawakami, and C. Mira, “Bifurcations of a quadratic two-dimensional diffeomorphism, homoclinic and heteroclinic situations”, *C. R. Acad. Sci., Paris* **298**, 253–256 (1984).
- [5] I. Gumowski and C. Mira, *Recurrences and Discrete Dynamical Systems* (Springer, New York, 1980).
- [6] K. T. Hansen, *Symbolic Dynamics in Chaotic systems*, PhD thesis (Univ. of Oslo, Oslo, Norway, 1993).
- [7] B.-L. Hao, *Elementary Symbolic Dynamics and Chaos in Dissipative Systems* (World Scientific, Singapore, 1989).
- [8] B.-L. Hao, *Chaos II* (World Scientific, Singapore, 1990).
- [9] M. Hénon, “A two-dimensional mapping with a strange attractor”, *Commun. Math. Phys.* **50**, 69 (1976).
- [10] M. Hénon, “On the numerical computation of Poincaré maps”, *Physica D* **5**, 412–414 (1982).

- [11] R. Lozi, “Un attracteur étrange (?) du type attracteur de Hénon”, *J. Phys. (Paris) Colloq.* **39**, C5–C9 (1978).
- [12] C. Mira, *Chaotic Dynamics - From One Dimensional Endomorphism to Two Dimensional Diffeomorphism* (World Scientific, Singapore, 1987).
- [13] M. Misiurewicz, “Strange attractors for the Lozi mapping”, *Ann. New York Acad. Sci.* **357**, 348–358 (1980).
- [14] S. Smale, “Differentiable dynamical systems”, *Bull. Amer. Math. Soc.* **73**, 747–817 (1967).
- [15] N. Tufillaro, T. Abbott, and J. Reilly, *An Experimental Approach to Non-linear Dynamics and Chaos* (Addison-Wesley, Redwood City, 1992).
- [16] L.-S. Young, “Bowen-Ruelle measures for certain piecewise hyperbolic maps”, *Trans. Amer. Math. Soc.* **287**, 41–48 (1985).

3.4 Examples

The reader is urged to study the examples collected at the ends of chapters. If you want to return back to the main text, click on [click to return] pointer on the margin.

What about smooth, continuous time flows, with no obvious surfaces that would be good Poincaré sections?

Example 3.1. Pendulum. The phase space of a simple pendulum is 2-dimensional: momentum on the vertical axis and position on the horizontal axis. We choose the Poincaré section to be the positive horizontal axis. Now imagine what happens as a point traces a trajectory through this phase space. As long as the motion is oscillatory, all pendulum orbits are loops. Thus, every trajectory periodically intersects the line at a single point, where the line represents the Poincaré section.

Consider next a pendulum with friction, such as the unforced Duffing system plotted in figure 2.4. Now every trajectory is an inward spiral, and the trajectory will intersect the Poincaré section $y = 0$ at a series of points that get closer and closer to either one of the equilibrium points, where the Duffing oscillator is at rest.

[return to p. 71](#)

Motion of a pendulum is so simple that you can sketch it yourself on a piece of paper. The next example (as well as example 30.4) offers a better illustration of the utility of visualization of dynamics by means of Poincaré sections.

Example 3.2. Rössler flow. (Continued from example 2.3) Consider figure 2.6, a typical trajectory of the 3-dimensional Rössler flow (2.30). The strange attractor wraps around the z axis, so one choice for a Poincaré section is a plane passing through the z axis. A sequence of such Poincaré sections placed radially at increasing angles with respect to the x axis, figure 3.3, illustrates the ‘stretch & fold’ action of the Rössler flow, by assembling these sections into a series of snapshots of the flow. A line segment in (a), traversing the width of the attractor at $y = 0, x > 0$ section, starts out close to the x - y plane, and after the stretching (a) $\rightarrow$ (b) followed by the folding (c) $\rightarrow$ (d), the folded segment returns (d) $\rightarrow$ (a) close to the initial segment, strongly compressed. In one Poincaré return the interval is thus stretched, folded and mapped onto itself, so the flow is *expanding*. It is also *mixing*, as demonstrated in one Poincaré return. In this situation, a point from the interior of the attractor maps onto the outer edge, while an edge point lands in the interior.

[exercise 3.1](#)

Once a particular Poincaré section is picked, we can also exhibit the return map (3.1), as in figure 3.4. Cases (a) and (d) are examples of nice 1-to-1 return maps. While (b) and (c) appear multimodal and non-invertible, these apparent traits are artifacts of projecting a 2-dimensional return map $(r_n, z_n) \rightarrow (r_{n+1}, z_{n+1})$ onto a 1-dimensional subspace $r_n \rightarrow r_{n+1}$. (continued in example 3.4)

[exercise 3.2](#)

[return to p. 71](#)

Example 3.3. Sections of Lorenz flow. (Continued from example 2.2) The plane $\mathcal{P}$ fixed by the $x = y$ diagonal and the z -axis depicted in figure 3.5 is a natural choice of a Poincaré section of the Lorenz flow of figure 2.5, as it contains all three equilibria, $x_{EQ_0} = (0, 0, 0)$ and the (2.26) pair x_{EQ_1}, x_{EQ_2} . This Poincaré section meets the orientation condition (3.4), and the points where the flow pierces *into* the section are marked by dots.



Equilibria x_{EQ_1} , x_{EQ_2} are centers of out-spirals, and close to them the section is transverse to the flow. However, close to EQ_0 , trajectories pass the z -axis either by crossing the section $\mathcal{P}$ or staying on the viewer's side. We are free to deploy as many sections as we wish; in order to capture the whole flow in this neighborhood, we add the second Poincaré section, $\mathcal{P}'$, through the $y = -x$ diagonal and the z -axis. Together the two sections, figure 3.5 (b), capture the whole flow near EQ_0 . In contrast to Rössler sections of figure 3.3, these appear very singular. We explain this singularity in example 4.6 and postpone construction of a return map until example 11.8. (E. Siminos and J. Halcrow)

[return to p. 74](#)

Example 3.4. Computation of Rössler flow Poincaré sections. (Continued from example 3.2) Convert Rössler equation (2.30) to cylindrical coordinates:

$$\begin{aligned}\dot{r} &= v_r = -z \cos \theta + ar \sin^2 \theta \\ \dot{\theta} &= v_\theta = 1 + \frac{z}{r} \sin \theta + \frac{a}{2} \sin 2\theta \\ \dot{z} &= v_z = b + z(r \cos \theta - c).\end{aligned}\tag{3.16}$$

Poincaré sections of figure 3.3 are defined by the fixing angle $U(x) = \theta - \theta_0 = 0$. In principle one should use the equilibrium x_+ from (2.31) as the origin, and its eigenvectors as the coordinate frame, but here original coordinates suffice. For parameter values (2.30) and (x_0, y_0, z_0) sufficiently far away from the inner equilibrium, θ increases monotonically with time. Integrate

$$\frac{dr}{d\theta} = v_r/v_\theta, \quad \frac{dt}{d\theta} = 1/v_\theta, \quad \frac{dz}{d\theta} = v_z/v_\theta\tag{3.17}$$

from (r_n, θ_n, z_n) to the next Poincaré section at θ_{n+1} , and switch the integration back to (x, y, z) coordinates. (continued in example 4.1) (Radford Mitchell, Jr.)

[return to p. 75](#)

Example 3.5. Hénon map. The map

$$\begin{aligned}x_{n+1} &= 1 - ax_n^2 + by_n \\ y_{n+1} &= x_n\end{aligned}\tag{3.18}$$

is a nonlinear 2-dimensional map frequently employed in testing various hunches about chaotic dynamics. The Hénon map is sometimes written as a 2-step recurrence relation

$$x_{n+1} = 1 - ax_n^2 + bx_{n-1}.\tag{3.19}$$

An n -step recurrence relation is the discrete-time analogue of an n th order differential equation, and it can always be replaced by a set of n 1-step recurrence relations.

The Hénon map is the simplest map that captures the ‘stretch & fold’ dynamics of return maps such as Rössler’s, figure 3.3. It can be obtained by a truncation of a polynomial approximation (3.15) to a return map (3.15) to second order.

A quick sketch of the long-time dynamics of such a mapping (an example is depicted in figure 3.7), is obtained by picking an arbitrary starting point and iterating (3.18) on a computer.

Always plot the dynamics of such maps in the (x_n, x_{n+1}) plane, rather than in the (x_n, y_n) plane, and make sure that the ordinate and abscissa scales are the same, so $x_n =$

x_{n+1} is the 45° diagonal. There are several reasons why one should plot this way: (a) we think of the Hénon map as a model return map $x_n \rightarrow x_{n+1}$, and (b) as parameter b varies, the attractor will change its y -axis scale, while in the (x_n, x_{n+1}) plane it goes to a parabola as $b \rightarrow 0$, as it should.

[exercise 3.5](#)

As we shall soon see, periodic orbits will be key to understanding the long-time dynamics, so we also plot a typical periodic orbit of such a system, in this case an unstable period 7 cycle. Numerical determination of such cycles will be explained in sect. 34.1, and the periodic point labels 0111010, 1110100, $\dots$ in sect. 15.2. (continued in example 15.4)

[return to p. 77](#)

Example 3.6. Lozi map. Another example frequently employed is the *Lozi map*, a linear, ‘tent map’ version of the Hénon map (3.18) given by

$$\begin{aligned} x_{n+1} &= 1 - a|x_n| + by_n \\ y_{n+1} &= x_n. \end{aligned} \quad (3.20)$$

Though not realistic as an approximation to a smooth flow, the Lozi map is a very helpful tool for developing intuition about the topology of a large class of maps of the ‘stretch & fold’ type.

[return to p. 77](#)

Example 3.7. Parabola. For a sufficiently large value of the stretching parameter a , one iteration of the Hénon map (3.18) stretches and folds a region of the (x, y) plane centered around the origin, as will be illustrated in figure 15.4. Parameter a controls the amount of stretching, while parameter b controls the thickness of the folded image through the ‘1-step memory’ term bx_{n-1} in (3.19). In figure 3.7, parameter b is rather large, $b = 0.3$, so the attractor is rather thick and the transverse fractal structure clearly visible. For vanishingly small b , the Hénon map reduces to a 1-dimensional quadratic map

$$x_{n+1} = 1 - ax_n^2. \quad (3.21)$$

By setting $b = 0$, we lose determinism: the inverse of map (3.21) has two real preimages $\{x_{n-1}^+, x_{n-1}^-\}$ for most x_n . If Bourbaki is your native dialect, the Hénon map is *injective* or one-to-one, but the quadratic map is *surjective* or many-to-one. Still, this 1-dimensional approximation is very instructive. (continued in example 14.6)

[exercise 3.6](#)

[return to p. 77](#)

Exercises

- 3.1. **Poincaré sections of the Rössler flow.** (continuation of exercise 2.8) Calculate numerically a Poincaré section (or several Poincaré sections) of the Rössler flow. Since Rössler flow state space is 3D, the flow maps onto a 2D Poincaré section. Do you see that in your numerical results? Does replacing the return map for this section by a 1-dimensional map give a good approximation? More precisely, estimate the thickness of the strange attractor. (continued as exercise 4.4)

(R. Paškauskas)

- 3.2. **A return return map for the Rössler flow.** (continuation of exercise 3.1) The return maps of figure 3.4 appear multimodal and non-invertible, which is an artifact of the 2-dimensional return map $(R_n, z_n) \rightarrow (R_{n+1}, z_{n+1})$ being projected onto the 1-dimensional subspace $R_n \rightarrow R_{n+1}$.

Construct a genuine $s_{n+1} = f(s_n)$ return map by parameterizing points on a Poincaré section of the attractor figure 3.3 by a Euclidean length s computed curvilinearly along the attractor section. (For a discussion of curvilinear parametrizations of invariant manifolds, see sect. 15.1.1.)

This is best done (using methods to be developed in what follows) by a continuation of the unstable manifold of the 1-cycle embedded in the strange attractor, figure 7.5 (b).

(P. Cvitanović)

- 3.3. **Arbitrary Poincaré sections.** We will generalize the construction of Poincaré sections, so that any section shape is allowable, as specified by the equation $U(x) = 0$.

- (a) Start by modifying your integrator so that you can change the coordinates once you get near the Poincaré section. You can do this easily by writing the equation as

$$\frac{dx_k}{ds} = \kappa f_k, \quad (3.22)$$

with $dt/ds = \kappa$, and choosing κ to be 1 or $1/f_1$. This allows one to switch between t and x_1 as the integration 'time.'

- (b) Introduce an extra dimension x_{n+1} into your system and set

$$x_{n+1} = U(x). \quad (3.23)$$

How can this be used to find a Poincaré section?

- 3.4. **Deterministic collinear helium dynamics.**

(continuation of exercise 2.11) Make a Poincaré section by plotting (r_1, p_1) whenever $r_2 = 0$: Note that for $r_2 = 0$, p_2 is already determined by (8.32). Compare your results with figure A2.3 (b).

(Gregor Tanner, Per Rosenqvist)

- 3.5. **Hénon map fixed points.** Show that the two fixed points $(x_0, x_0), (x_1, x_1)$ of the Hénon map (3.18) are given by

$$\begin{aligned} x_0 &= \frac{-(1-b) - \sqrt{(1-b)^2 + 4a}}{2a}, \\ x_1 &= \frac{-(1-b) + \sqrt{(1-b)^2 + 4a}}{2a}. \end{aligned} \quad (3.24)$$

- 3.6. **Fixed points of maps.** A continuous function F is a contraction of the unit interval if it maps the interval inside itself.
- (a) Use the continuity of F to show that a 1-dimensional contraction F of the interval $[0, 1]$ has at least one fixed point.
 - (b) In a uniform (hyperbolic) contraction, the slope of F is always smaller than one, $|F'| < 1$. Is the composition of uniform contractions a contraction? Is it uniform?
- 3.7. **Section border for Rössler.** (continuation of exercise 3.1) Determine numerically section borders (3.8) for several Rössler flow Poincaré sections of exercise 3.1 and figure 3.3, at least for angles
- (a) -60° , (b) 0° , and
 - (c) A Poincaré section hyperplane that goes through both equilibria, see (2.31) and figure 4.5. Two points only fix a line: think of a criterion for a good orientation of the section hyperplane, perhaps by demanding that the contracting eigenvector of the 'inner' equilibrium x_- lies in it.
 - (d) (Optional) Hand- or computer-draw a visualization of the section border as a 3-dimensional fluid flow, which either crosses, is tangent to, or fails to cross a sheet of light cutting across the flow.

As the state space is 3-dimensional, the section borders are 1-dimensional, and it should be easy to outline the border by plotting the color-coded magnitude of $v_\perp(\hat{x})$, the component of $v(\hat{x})$ normal to the section, for a fine grid of 2-dimensional Poincaré section plane points. For sections that go through the z -axis, the normal velocity $v_\perp(\hat{x})$ is tangent to the circle through $\hat{x}$, and vanishes for θ in polar coordinates (3.16), but that is not true for other Poincaré sections, such as in case (c).

(P. Cvitanović)

Chapter 4

Local stability

It does not say in the Bible that all laws of nature are expressible linearly.

— Enrico Fermi

(R. Mainieri and P. Cvitanović)

SO FAR we have concentrated on describing the trajectory of a single initial point. Our next task is to define and determine the size of a *neighborhood* of $x(t)$. We shall do this by assuming that the flow is locally smooth and by describing the local geometry of the neighborhood by studying the flow linearized around $x(t)$. Nearby points aligned along the stable (contracting) directions remain in the neighborhood of the trajectory $x(t) = f^t(x_0)$; the ones to keep an eye on are the points which leave the neighborhood along the unstable directions. As we shall demonstrate in chapter 21, the expanding directions matter in hyperbolic systems. The repercussions are far-reaching. As long as the number of unstable directions is finite, the same theory applies to finite-dimensional ODEs, state space volume preserving Hamiltonian flows, and dissipative, volume contracting infinite-dimensional PDEs.



In order to streamline the exposition, in this chapter all examples are collected in sect. 4.8. We strongly recommend that you work through these examples. You can navigate to these examples and back by clicking on the [example] links, such as



example 4.8
p. 101

4.1 Flows transport neighborhoods



As a swarm of representative points moves along, it carries along and distorts neighborhoods. The deformation of an infinitesimal neighborhood is best un-



derstood by considering a trajectory originating near $x_0 = x(0)$, with an initial infinitesimal deviation vector $\delta x(0)$. The flow then transports the deviation vector $\delta x(t)$ along the trajectory $x(x_0, t) = f^t(x_0)$.

4.1.1 Instantaneous rate of shear

The system of linear *equations of variations* for the displacement of the infinitesimally close neighbor $x + \delta x$ follows from the flow equations (2.9) by Taylor expanding to linear order

$$\dot{x}_i + \dot{\delta x}_i = v_i(x + \delta x) \approx v_i(x) + \sum_j \frac{\partial v_i}{\partial x_j} \delta x_j.$$

The infinitesimal deviation vector δx is thus transported along the trajectory $x(x_0, t)$, with time variation given by

$$\frac{d}{dt} \delta x_i(x_0, t) = \sum_j \left. \frac{\partial v_i}{\partial x_j}(x) \right|_{x=x(x_0, t)} \delta x_j(x_0, t). \quad (4.1)$$

As both the displacement and the trajectory depend on the initial point x_0 and the time t , we shall often abbreviate the notation to $x(x_0, t) \rightarrow x(t) \rightarrow x$, $\delta x_i(x_0, t) \rightarrow \delta x_i(t) \rightarrow \delta x$ in what follows. Taken together, the set of equations

$$\dot{x}_i = v_i(x), \quad \dot{\delta x}_i = \sum_j A_{ij}(x) \delta x_j \quad (4.2)$$

governs the dynamics in the tangent bundle $(x, \delta x) \in \mathbf{TM}$ obtained by adjoining the d -dimensional tangent space $\delta x \in T\mathcal{M}_x$ to every point $x \in \mathcal{M}$ in the d -dimensional state space $\mathcal{M} \subset \mathbb{R}^d$. The *stability matrix* or *velocity gradients matrix*

remark 4.1

$$A_{ij}(x) = \frac{\partial}{\partial x_j} v_i(x) \quad (4.3)$$

describes the instantaneous rate of shearing of the infinitesimal neighborhood of $x(t)$ by the flow. A swarm of neighboring points of $x(t)$ is instantaneously sheared by the action of the stability matrix, $\delta x(t + \delta t) = \delta x(t) + \delta t A(x_n) \delta x(t)$. A is a tensorial rate of deformation, so it is a bit hard (if not impossible) to draw.



example 4.1
p. 101

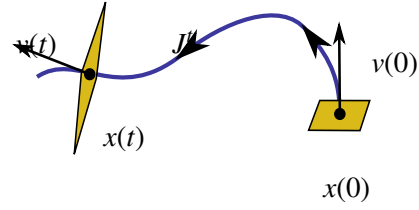
4.1.2 Finite time linearized flow

By Taylor expanding a *finite time* flow to linear order,



$$f_i^t(x_0 + \delta x) = f_i^t(x_0) + \sum_j \frac{\partial f_i^t(x_0)}{\partial x_{0j}} \delta x_j + \dots, \quad (4.4)$$

Figure 4.1: For finite times a local frame is transported along the orbit and deformed by Jacobian matrix J^t . As J^t is not self-adjoint, an initial orthogonal frame is mapped into a non-orthogonal one.



one finds that the linearized neighborhood is transported by the Jacobian matrix

remark 4.1

$$\delta x(t) = J^t(x_0) \delta x_0, \quad J_{ij}^t(x_0) = \frac{\partial x(t)_i}{\partial x(0)_j}, \quad J^0(x_0) = \mathbf{1}. \quad (4.5)$$

For example, in 2 dimensions the Jacobian matrix for change from initial to final coordinates is

$$J^t = \frac{\partial(x, y)}{\partial(x_0, y_0)} = \begin{bmatrix} \frac{\partial x}{\partial x_0} & \frac{\partial x}{\partial y_0} \\ \frac{\partial y}{\partial x_0} & \frac{\partial y}{\partial y_0} \end{bmatrix}.$$

The Jacobian matrix is evaluated on a trajectory segment that starts at point $x_0 = x(t_0)$ and ends at point $x_1 = x(t_1)$, $t_1 \geq t_0$. As the trajectory $x(t)$ is deterministic, the initial point x_0 and the elapsed time t in (4.5) suffice to determine J , but occasionally we find it helpful to be explicit about the initial and final times and state space positions, and write

$$J_{ij}^{t_1-t_0} = J_{ij}(t_1; t_0) = J_{ij}(x_1, t_1; x_0, t_0) = \frac{\partial x(t_1)_i}{\partial x(t_0)_j}. \quad (4.6)$$

question 4.1

The map f^t is assumed invertible and differentiable so that J^t exists. For sufficiently short times J^t remains close to $\mathbf{1}$, so $\det J^t > 0$. By continuity, $\det J^t$ remains positive for all times t . However, for discrete time maps, $\det J^n$ can have either sign.

4.1.3 Co-moving frames

J describes the deformation of an infinitesimal neighborhood at a finite time t in the co-moving frame of $x(t)$. This deformation of an initial frame at x_0 into a non-orthogonal frame at $x(t)$ is described by the eigenvectors and eigenvalues of the Jacobian matrix of the linearized flow (see figure 4.1),

$$J^t \mathbf{e}^{(j)} = \Lambda_j \mathbf{e}^{(j)}, \quad j = 1, 2, \dots, d. \quad (4.7)$$

Throughout this text the symbol Λ_k will always denote the k th *eigenvalue* (the *stability multiplier*) of the finite time Jacobian matrix J^t . Symbol $\lambda^{(k)}$ will be reserved for the k th *stability exponent*, with real part $\mu^{(k)}$ and phase $\omega^{(k)}$:

$$\Lambda_k = e^{t\lambda^{(k)}} \quad \lambda^{(k)} = \mu^{(k)} + i\omega^{(k)}. \quad (4.8)$$

As J^t is a real matrix, its eigenvalues are either real or come in complex conjugate pairs,

$$\{\Lambda_k, \Lambda_{k+1}\} = \{e^{t(\mu^{(k)} + i\omega^{(k)})}, e^{t(\mu^{(k)} - i\omega^{(k)})}\},$$

with magnitude $|\Lambda_k| = |\Lambda_{k+1}| = \exp(t\mu^{(k)})$. The phase $\omega^{(k)}$ describes the rotation velocity in the plane spanned by the pair of real eigenvectors, $\{\text{Re } \mathbf{e}^{(k)}, \text{Im } \mathbf{e}^{(k)}\}$, with one period of rotation given by $T = 2\pi/\omega^{(k)}$.



example 4.4
p. 102

$J^t(x_0)$ depends on the initial point x_0 and the elapsed time t . For notational brevity we omitted this dependence, but in general both the eigenvalues and the eigenvectors, $\Lambda_j = \Lambda_j(x_0, t)$, $\dots$, $\mathbf{e}^{(j)} = \mathbf{e}^{(j)}(x_0, t)$, also depend on the trajectory traversed.



Nearby trajectories diverge exponentially with time along *unstable directions* and approach each other along *stable directions*; however, the distance between trajectories both increases and decreases with time along *marginal directions* at rates slower than exponential. The relative path of nearby trajectories (i.e., diverging, approaching or changing) corresponds to the multipliers (eigenvalues) of the Jacobian matrix with magnitude larger than, smaller than, or equal to 1. In the literature, the adjectives *neutral*, *indifferent* and *center* are often used instead of ‘marginal’. Attracting, or stable directions are sometimes called ‘asymptotically stable’, and so on.

One of the preferred directions is what one might expect, which is the direction of the flow itself. To see that, consider two initial points along a trajectory separated by infinitesimal flight time δt : $\delta x_0 = f^{\delta t}(x_0) - x_0 = v(x_0)\delta t$. By the semigroup property of the flow, $f^{t+\delta t} = f^{\delta t} \circ f^t$, where



$$f^{\delta t+t}(x_0) = \int_t^{\delta t+t} d\tau v(x(\tau)) + f^t(x_0) = \delta t v(x(t)) + f^t(x_0).$$

Expanding both sides of $f^t(f^{\delta t}(x_0)) = f^{\delta t}(f^t(x_0))$, keeping the leading term in δt , and using the definition of the Jacobian matrix (4.5), we observe that $J^t(x_0)$ transports the velocity vector at x_0 to the velocity vector at $x(t)$ (see figure 4.1):

$$v(x(t)) = J^t(x_0) v(x_0). \quad (4.9)$$

4.2 Computing the Jacobian matrix

We started by assuming that we know the equations of motion, and from (4.3) we also know stability matrix A , the instantaneous rate of shear of an infinitesimal neighborhood $\delta x_i(t)$ of the trajectory $x(t)$. What we do not know is the finite time deformation (4.5), so our next task is to relate the stability matrix A to the Jacobian matrix J^t . In terms of differential equations, the relationship between these two matrices is found by taking the time derivative of (4.5) and replacing δx by (4.2)



$$\frac{d}{dt} \delta x(t) = \frac{dJ^t}{dt} \delta x_0 = A \delta x(t) = AJ^t \delta x_0.$$

Hence the matrix elements of the $[d \times d]$ Jacobian matrix satisfy the ‘tangent linear equations’

$$\frac{d}{dt} J^t(x_0) = A(x) J^t(x_0), \quad x = f^t(x_0), \quad \text{initial condition } J^0(x_0) = \mathbf{1}. \quad (4.10)$$

For autonomous flows, the matrix of velocity gradients $A(x)$ depends only on x , not time, while J^t depends on both the state space position and time. Given a numerical routine for integrating the equations of motion, evaluation of the Jacobian matrix requires minimal additional programming effort; one simply extends the d -dimensional integration routine and integrates the d^2 elements of $J^t(x_0)$ concurrently with $f^t(x_0)$.

The qualifier ‘simply’, however, is simply a scam perpetrated by introductory ‘chaos’ courses. (1) Integration will work for short finite times, but for exponentially unstable flows one quickly runs into numerical over- and/or underflow problems. (2) For high-dimensional flows the analytical expressions for elements of A are probably so cumbersome that A fits on no computer. Further thought will have to go into implementing this calculation.

chapter 30

So now we know how to compute Jacobian matrix J^t given the stability matrix A , at least when the d^2 extra equations are not too expensive to compute. Mission accomplished.



fast track:
chapter 8, p. 146

And yet... there are mopping up operations left to do. We persist until we derive the integral formula (4.19) for the Jacobian matrix, an analogue of the finite-time ‘Green’s function’ or ‘path integral’ solutions of other linear problems.

We are interested in smooth, differentiable flows. If a flow is smooth, in a sufficiently small neighborhood it is essentially linear. Hence the next section, which might seem an embarrassment (what is a section on *linear* flows doing in a book on *nonlinear* dynamics?), offers a firm stepping stone on the way to understanding nonlinear flows. Linear charts are the key tool of differential geometry, general relativity, etc., so we are in good company. If you know your eigenvalues and eigenvectors, you may prefer to fast forward here.



fast track:
sect. 4.4, p. 92

4.3 A linear diversion

Linear is good, nonlinear is bad.
—Jean Bellissard

Linear fields are the simplest vector fields, described by linear differential equations which can be solved explicitly, with solutions that are good for all times. The state space for linear differential equations is $\mathcal{M} = \mathbb{R}^d$, and the equations of motion (2.9) are written in terms of a vector x and a constant stability matrix A as

$$\dot{x} = v(x) = Ax. \quad (4.11)$$

Solving this equation means finding the state space trajectory

$$x(t) = (x_1(t), x_2(t), \dots, x_d(t))$$

passing through a given initial point x_0 . If $x(t)$ is a solution with $x(0) = x_0$ and $y(t)$ another solution with $y(0) = y_0$, then the linear combination $ax(t) + by(t)$ with $a, b \in \mathbb{R}$ is also a solution, but now starting at the point $ax_0 + by_0$. At any instant in time, the space of solutions is a d -dimensional vector space, spanned by a basis of d linearly independent solutions.

How do we solve the linear differential equation (4.11)? If instead of a matrix equation we have a scalar one, $\dot{x} = \lambda x$, the solution is $x(t) = e^{t\lambda}x_0$. In order to solve the d -dimensional matrix case, it is helpful to rederive this solution by studying what happens for a short time step δt . If time $t = 0$ coincides with position $x(0)$, then

$$\frac{x(\delta t) - x(0)}{\delta t} = \lambda x(0), \quad (4.12)$$

which we iterate m times to obtain Euler's formula for compounding interest

$$x(t) \approx \left(1 + \frac{t}{m}\lambda\right)^m x(0) \approx e^{t\lambda}x(0). \quad (4.13)$$

The term in parentheses acts on the initial condition $x(0)$ and evolves it to $x(t)$ by taking m small time steps $\delta t = t/m$. As $m \rightarrow \infty$, the term in parentheses converges to $e^{t\lambda}$. Consider now the matrix version of equation (4.12):

$$\frac{x(\delta t) - x(0)}{\delta t} = Ax(0). \quad (4.14)$$

A representative point x is now a vector in $\mathbb{R}^d$ acted on by the matrix A , as in (4.11). By denoting the identity matrix as $\mathbf{1}$ and repeating steps (4.12) and (4.13), we obtain Euler's formula for the exponential of a matrix:

$$x(t) = J^t x(0), \quad J^t = e^{tA} = \lim_{m \rightarrow \infty} \left(\mathbf{1} + \frac{t}{m}A\right)^m, \quad (4.15)$$

where $J^t = J(t)$ is a short hand for $\exp(tA)$. We will find this definition for the exponential of a matrix helpful in the general case, where the matrix $A = A(x(t))$ varies along a trajectory.



example 4.2
p. 101



example 4.3
p. 101

Now that we have some feeling for the qualitative behavior of eigenvectors and eigenvalues of linear flows, we are ready to return to the nonlinear case. What replaces the exponential in (4.15) then?

question 4.3
section 5.2.1

4.4 Stability of flows



How does one determine the eigenvalues of the finite time local deformation J^t for a general nonlinear smooth flow? The Jacobian matrix is computed by integrating the equations of variations (4.2)

$$x(t) = f^t(x_0), \quad \delta x(x_0, t) = J^t(x_0) \delta x(x_0, 0). \quad (4.16)$$

The equations for $J^t(x_0)$ are linear, so we should be able to integrate them—but in order to make sense of the answer, we derive this integral step by step.

Consider the case of a general, non-stationary trajectory $x(t)$. The exponential of a constant matrix can be defined either by its Taylor series expansion or in terms of the Euler limit (4.15):

$$e^{tA} = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k = \lim_{m \rightarrow \infty} \left(\mathbf{1} + \frac{t}{m} A \right)^m. \quad (4.17)$$

Taylor expanding is fine if A is a constant matrix. However, only the second, tax-accountant's discrete step definition of an exponential is appropriate for the task at hand. For dynamical systems, the local rate of neighborhood distortion $A(x)$ depends on where we are along the trajectory. The linearized neighborhood is deformed along the flow, and the m discrete time-step approximation to J^t is therefore given by a generalization of the Euler product (4.17):

$$\begin{aligned} J^t(x_0) &= \lim_{m \rightarrow \infty} \prod_{n=m}^1 \left(\mathbf{1} + \delta t A(x_n) \right) = \lim_{m \rightarrow \infty} \prod_{n=m}^1 e^{\delta t A(x_n)} \\ &= \lim_{m \rightarrow \infty} e^{\delta t A(x_m)} e^{\delta t A(x_{m-1})} \dots e^{\delta t A(x_2)} e^{\delta t A(x_1)}, \end{aligned} \quad (4.18)$$

where $\delta t = (t - t_0)/m$, and $x_n = x(t_0 + n\delta t)$. “Reverse” order in the product indices is to remind you that the successive infinitesimal deformation are applied by multiplying from the left. The $m \rightarrow \infty$ limit of this procedure is the formal integral that you might have seen in quantum mechanics courses, but now turns out to have nothing to do with ‘quantum’, it’s just linear algebra,

$$J_{ij}^t(x_0) = \left[\mathbf{T} e^{\int_0^t d\tau A(x(\tau))} \right]_{ij}, \quad (4.19)$$

where $\mathbf{T}$ stands for time-ordered integration, *defined* as the continuum limit of successive multiplications (4.18). This integral formula for J^t is the main conceptual result of the present chapter. It is the finite time companion of the differential definition (4.10). An important property of Jacobian matrices is evident from (4.18): these matrices are multiplicative along the direction of the flow,

$$J^{t+t'}(x) = J^{t'}(x') J^t(x), \quad \text{where } x' = f^{t'}(x_0). \quad (4.20)$$

This is an immediate consequence of the time-ordered product structure of (4.18). However, in practice J is evaluated by integrating (4.10) along with the ODEs that define a particular flow.



example 4.5
p. 103



example 4.6
p. 104



example 4.7
p. 105



exercise 4.5

4.5 Stability of maps



The transformation of an infinitesimal neighborhood of a trajectory under the iteration of a map follows from Taylor expanding the iterated mapping at finite time n to linear order, as in (4.4). The linearized neighborhood is transported by the Jacobian matrix evaluated at a discrete set of times $n = 1, 2, \dots$,



$$J_{ij}^n(x_0) = \left. \frac{\partial f_i^n(x)}{\partial x_j} \right|_{x=x_0}. \quad (4.21)$$

As in the finite time case (4.8), Λ_k denotes the k th *eigenvalue* or multiplier of the finite time Jacobian matrix J^n . There is really no difference from the continuous time case, other than that now the Jacobian matrix is evaluated at integer times.



example 4.9
p. 106

The formula for the linearization of n th iterate of a d -dimensional map

$$J^n(x_0) = J(x_{n-1}) \cdots J(x_1)J(x_0), \quad x_j = f^j(x_0), \quad (4.22)$$

in terms of single time steps $J_{jl} = \partial f_j / \partial x_l$ follows from the chain rule for functional composition,

$$\frac{\partial}{\partial x_i} f_j(f(x)) = \sum_{k=1}^d \left. \frac{\partial f_j(y)}{\partial y_k} \right|_{y=f(x)} \frac{\partial f_k(x)}{\partial x_i}.$$

If you prefer to think of a discrete time dynamics as a sequence of Poincaré section returns, then (4.22) follows from (4.20); Jacobian matrices are multiplicative along the flow.

exercise 6.3



example 4.10
p. 106



fast track:
chapter 8, p. 146

4.6 Stability of return maps



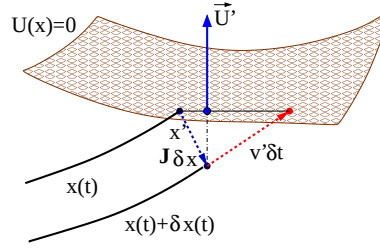
(R. Paškauskas and P. Cvitanović)

We now relate the linear stability of the return map $P : \mathcal{P} \rightarrow \mathcal{P}$ defined in sect. 3.1 to the stability of the continuous time flow in the full state space.



The hypersurface $\mathcal{P}$ can be specified implicitly through a function $U(x)$ that is zero whenever a point x is on the Poincaré section. A nearby point $x + \delta x$ is in the hypersurface $\mathcal{P}$ if $U(x + \delta x) = 0$, and the same is true for variations around the first return point $x' = x(\tau)$. Thus, expanding $U(x')$ to linear order in

Figure 4.2: If $x(t)$ intersects the Poincaré section $\mathcal{P}$ at time τ , the nearby $x(t) + \delta x(t)$ trajectory intersects it time $\tau + \delta t$ later. As $(U' \cdot v' \delta t) = -(U' \cdot J \delta x)$, the difference in arrival times is given by $\delta t = -(U' \cdot J \delta x) / (U' \cdot v')$.



variation δx restricted to the Poincaré section, and applying the chain rule leads to the condition

$$\sum_{i=1}^d \frac{\partial U(x')}{\partial x_i} \frac{dx'_i}{dx_j} \Big|_{\mathcal{P}} = 0. \quad (4.23)$$

In what follows $U_i = \partial_i U$ is the gradient of U defined in (3.3), unprimed quantities refer to the starting point $x = x_0 \in \mathcal{P}$, $v = v(x_0)$, and the primed quantities to the first return: $x' = x(\tau)$, $v' = v(x')$, $U' = U(x')$. For brevity we shall also denote the full state space Jacobian matrix at the first return by $J = J^\tau(x_0)$. Both the first return x' and the time of flight to the next Poincaré section $\tau(x)$ depend on the starting point x , so the Jacobian matrix

$$\hat{J}(x)_{ij} = \frac{dx'_i}{dx_j} \Big|_{\mathcal{P}} \quad (4.24)$$

with both initial and the final variation constrained to the Poincaré section hypersurface $\mathcal{P}$ is related to the continuous flow Jacobian matrix by

$$\frac{dx'_i}{dx_j} \Big|_{\mathcal{P}} = \frac{\partial x'_i}{\partial x_j} + \frac{dx'_i}{d\tau} \frac{d\tau}{dx_j} = J_{ij} + v'_i \frac{d\tau}{dx_j}.$$

The return time variation $d\tau/dx$, figure 4.2, is eliminated by substituting this expression into the constraint (4.23),

$$0 = \partial_i U' J_{ij} + (v' \cdot \partial U') \frac{d\tau}{dx_j},$$

yielding the projection of the full space d -dimensional Jacobian matrix to the return map $(d-1)$ -dimensional Jacobian matrix:

$$\hat{J}_{ij} = \left(\delta_{ik} - \frac{v'_i \partial_k U'}{v' \cdot \partial U'} \right) J_{kj}. \quad (4.25)$$

Substituting (4.9) we verify that the initial velocity $v(x)$ is a zero-eigenvector of $\hat{J}$

$$\hat{J}v = 0, \quad (4.26)$$

so the Poincaré section eliminates variations parallel to v , and $\hat{J}$ is a rank $(d-1)$ -dimensional matrix, i.e., one less than the dimension of the continuous time flow.

4.7 Neighborhood volume



section 6.2
remark 6.1

Consider an infinitesimal state space volume $\Delta V = dx_1 dx_2 \cdots dx_d$ centered around the point x_0 at time $t = 0$. The volume $\Delta V'$ around the point $x' = x(t)$ time t later is (see figure 4.1)

$$\Delta V' = \frac{\Delta V'}{\Delta V} \Delta V = \left| \det \frac{\partial x'}{\partial x} \right| \Delta V = \left| \det J^t(x_0) \right| \Delta V, \quad (4.27)$$

so the $|\det J|$ is the ratio of the initial and final volumes. The determinant $\det J^t(x_0) = \prod_{i=1}^d \Lambda_i(x_0, t)$ is the product of the Jacobian matrix (4.5) stability multipliers (4.8). We shall refer to this determinant as the *Jacobian* of the flow. To evaluate it, use the matrix identity $\ln \det J = \text{tr} \ln J$, take the time derivative and substitute the J evolution equation (4.10):

exercise 4.1

$$\frac{d}{dt} \ln \Delta V(t) = \frac{d}{dt} \ln \det J = \text{tr} \frac{d}{dt} \ln J = \text{tr} \left(\frac{1}{J} \frac{dJ}{dt} \right) = \text{tr} A = \partial_i v_i.$$

(Here, as elsewhere in ChaosBook, a repeated index implies summation.) Integrate both sides to obtain the time evolution of an infinitesimal volume (see [Liouville's formula wiki](#))

$$\det J^t(x_0) = \exp \left[\int_0^t d\tau \text{tr} \mathbf{A}(x(\tau)) \right] = \exp \left[\int_0^t d\tau \partial_i v_i(x(\tau)) \right]. \quad (4.28)$$

The divergence $\partial_i v_i$ characterizes the behavior of a state space volume in the infinitesimal neighborhood of the trajectory. As this is a scalar quantity, the integral in the exponent (4.19) needs *no time ordering*. So all we need to do is evaluate the time average

$$\begin{aligned} \overline{\partial_i v_i} &= \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t d\tau \sum_{i=1}^d A_{ii}(x(\tau)) \\ &= \lim_{t \rightarrow \infty} \frac{1}{t} \ln \left| \prod_{i=1}^d \Lambda_i(x_0, t) \right| = \sum_{i=1}^d \mu^{(i)}(x_0, \infty) \end{aligned} \quad (4.29)$$

along the trajectory. If the flow is not singular, the stability matrix elements are everywhere bounded from above, $A_{ij} < M$, and so is the trace $\sum_i A_{ii}$. The time integral in (4.28) thus grows at most linearly with t , when the divergence $\partial_i v_i$ is bounded for all times, and numerical estimates of the $t \rightarrow \infty$ limit in (4.29) are not marred by any blowups. In numerical evaluations of stability exponents, the sum rule (4.29) can serve as a helpful check on the accuracy of the computation.



example 4.8
p. 105

If $\partial_i v_i < 0$ at a given state space point x , the flow is *locally* contracting, and the trajectory might be falling into an attractor. If $\partial_i v_i(x) < 0$ for all $x \in \mathcal{M}$, the flow is *globally* contracting, with the dimension of the attractor necessarily smaller

than the dimension of state space $\mathcal{M}$. For ∞ -dimensional dissipative flows, such as Navier-Stokes, the ∞ of stability exponents $\mu^{(i)}$ in (4.29) can be arbitrarily negative; as such exponents represent damping of arbitrarily kinky modes of a viscous fluid, and they are of no interest for studying steady turbulence. So the sum (4.29) should be truncated to a finite number d_{phys} of leading stability exponents. We refer to the integer d_{phys} as the *physical* dimension of a *strange attractor*; however, in fluid dynamics it is often referred to as the *inertial manifold*. Every expanding or marginal direction contributes 1 to d_{phys} , and then to get a lower bound on d_{phys} , one has to keep at least as many negative $\mu^{(i)}$ as needed to ensure that the sum (4.29) is globally contracting. Since nonlinear terms can mix various terms in such a way that expansion in some directions overwhelms the strongly contracting ones, d_{phys} is larger than this bound, but still a finite number.

This is an amazing result: a fluid's state space is ∞ -dimensional, but its long term dynamics is confined to a finite-dimensional(!) subspace. This is the reason we can apply the few degrees of freedom technology developed here to ∞ -dimensional field theories.

If $\partial_i v_i = 0$, the flow preserves state space volume, $\det J^t = 1$, and the flow is *incompressible*. An important class of such flows are the Hamiltonian flows considered in chapter 8. But before we can get to that, Alice, the perfect student and always alert, pipes up, and asks question 4.4.

question 4.4

Résumé

A neighborhood of a trajectory deforms as it is transported by a flow. Let us summarize the linearized flow notation used throughout ChaosBook.

Differential formulation, flows: Equations

$$\dot{x} = v, \quad \delta\dot{x} = A \delta x$$

govern the dynamics in the tangent bundle $(x, \delta x) \in \mathbf{TM}$ obtained by adjoining the d -dimensional tangent space $\delta x \in T\mathcal{M}_x$ to every point $x \in \mathcal{M}$ in the d -dimensional state space $\mathcal{M} \subset \mathbb{R}^d$. In the linear approximation, the stability matrix $A = \partial v / \partial x$ describes the instantaneous rate of shearing / compression / expansion of an infinitesimal neighborhood of state space point x .

Finite time formulation, maps: A discrete set of trajectory points $\{x_0, x_1, \dots, x_n, \dots\} \in \mathcal{M}$ can be generated by composing finite-time maps, either given as $x_{n+1} = f(x_n)$, or obtained by integrating the dynamical equations

$$x_{n+1} = f^{\Delta t_n}(x_n) = x_n + \int_{t_n}^{t_{n+1}} d\tau v(x(\tau)), \quad \Delta t_n = t_{n+1} - t_n, \quad (4.30)$$

for a discrete sequence of times $\{t_0, t_1, \dots, t_n, \dots\}$, specified by some criterion such as strobing or Poincaré sections. In the discrete time formulation the dynamics in the tangent bundle $(x, \delta x) \in \mathbf{TM}$ is governed by

$$x_{n+1} = f(x_n), \quad \delta x_{n+1} = J(x_n) \delta x_n,$$

where

$$J(x_n) = J^{\Delta t_n}(x_n) = \frac{\partial x_{n+1}}{\partial x_n}$$

is the 1-time step Jacobian matrix. The deformation after a finite time t is described by the Jacobian matrix

$$J^t(x_0) = \mathbf{T} e^{\int_0^t d\tau A(x(\tau))},$$

where $\mathbf{T}$ stands for the time-ordered integration, defined multiplicatively along the trajectory. For discrete time maps, this represents multiplication by the time-step Jacobian matrix J along the n points $x_0, x_1, x_2, \dots, x_{n-1}$ on the trajectory of x_0 ,

$$J^n(x_0) = J(x_{n-1}) J(x_{n-2}) \cdots J(x_1) J(x_0),$$

where $J(x)$ is the 1-time step Jacobian matrix.

In ChaosBook the *stability multiplier* Λ_k denotes the k th *eigenvalue* of the finite time Jacobian matrix $J^t(x_0)$, $\mu^{(k)}$ the real part of k th *stability exponent*, and $\theta^{(k)}$ its phase,

$$\Lambda = e^{t\mu + i\theta}.$$

For complex eigenvalue pairs the ‘angular velocity’ ω describes rotational motion in the plane spanned by the real and imaginary parts of the corresponding pair of complex eigenvectors. This angular velocity ω has to be carefully “unwrapped” because most numerical routines return

$$\theta = t\omega \quad \text{mod } 2\pi.$$

The eigenvalues and eigen-directions of the Jacobian matrix describe the deformation of an initial infinitesimal cloud of neighboring trajectories into a distorted cloud at a finite time t later. Nearby trajectories separate exponentially along unstable eigen-directions, approach each other along stable directions, and the distance between trajectories changes slowly along marginal or center directions. The Jacobian matrix J^t is in general neither symmetric, nor diagonalizable by a rotation, nor do its (left or right) eigenvectors define an orthonormal coordinate frame. Furthermore, although the Jacobian matrices are multiplicative along the flow, their eigenvalues are generally not multiplicative in dimensions higher than one. This lack of a multiplicative nature for eigenvalues has important repercussions for both deterministic and quantum dynamics.

Commentary

Remark 4.1. Linear flows. The subject of linear algebra generates innumerable tomes of its own; in sect. 4.3 we only sketch, and in appendix A4 recapitulate a few facts that our narrative relies on. A useful reference book is Meyer [16]. The basic facts are presented at length in many textbooks. Frequently cited linear algebra references are Golub

and Van Loan [7], Coleman and Van Loan [4], and Watkins [24, 25]. The standard references that exhaustively enumerate and explain all possible cases are Hirsch and Smale [9] and Arnol'd [2]. A quick overview is given by Izhikevich [11]; for different notions of orbit stability see Holmes and Shea-Brown [10]. For ChaosBook purposes, we enjoyed the discussion in chapter 2 Meiss [15], chapter 1 of Perko [17] and chapters 3 and 5 of Glendinning [5]; we also liked the discussion of norms, least square problems, and differences between singular value and eigenvalue decompositions in Trefethen and Bau [22]. Appendix A of Stone and Goldbart [20] is an advanced summary of almost everything a graduate student needs to know about linear algebra. More pedestrian and perhaps easier to read is Chapter 3 of Arfken and Weber [1]. Truesdell [23] and Gurtin [8] are excellent references for the continuum mechanics perspective on state space dynamics. For a gentle introduction into parallels between dynamical systems and continuum mechanics, see Christov *et al.* [3].

section 6.1

The nomenclature tends to be a bit confusing. A Jacobian matrix (4.5) is sometimes referred to as the *fundamental solution matrix* or simply *fundamental matrix*, a name inherited from the theory of linear ODEs, or the *Fréchet derivative* of the nonlinear mapping $f'(x)$, or the *'tangent linear propagator'*, or even as the 'error matrix' (Lorenz [13]). The formula (4.22) for the linearization of n th iterate of a d -dimensional map is called a *linear cocycle*, a *multiplicative cocycle*, a *derivative cocycle* or simply a *cocycle* by some. Since matrix J describes the deformation of an infinitesimal neighborhood at a finite time t in the co-moving frame of $x(t)$, it is called a *deformation gradient* or a *transplacement gradient* in continuum mechanics. It is often denoted Df , but for our needs (we shall have to sort through a plethora of related Jacobian matrices) matrix notation J is more economical. Single discrete time-step Jacobian $J_{jl} = \partial f_j / \partial x_l$ in (4.22) is referred to as the 'tangent map' by Skokos [18, 19]. For a discussion of 'fundamental matrix' see appendix A4.2.

We follow Tabor [21] in referring to A in (4.3) as the 'stability matrix'; it is also referred to as the 'velocity gradient matrix' or 'velocity gradient tensor'. Matrix A is used to describe stability of equilibria, time-invariant points in state space; whereas, the stability of trajectories is described by Jacobian matrices. Goldhirsch, Sulem, and Orszag [6] call it the 'Hessenberg matrix', and refer to the equations of variations (4.1) as 'stability equations.' Manos *et al.* [14] refer to (4.1) as the 'variational equations'.

Sometimes A , which describes the instantaneous shear of the neighborhood of $x(x_0, t)$, is referred to as the 'Jacobian matrix', a particularly unfortunate usage when one considers linearized stability of an equilibrium point (5.1). A is not a Jacobian matrix, just as a generator of $SO(2)$ rotation is not a rotation; A is a generator of an infinitesimal time step deformation, $J^{\delta t} \simeq \mathbf{1} + A\delta t$. What Jacobi had in mind in his 1841 fundamental paper [12] on determinants (today known as 'Jacobians') were transformations between different coordinate frames. These are dimensionless quantities, while dimensionally A_{ij} is $1/[\text{time}]$.

More unfortunate still is referring to the Jacobian matrix $J^t = \exp(tA)$ as an 'evolution operator', which here (see sect. 20.2) refers to something altogether different. In ChaosBook Jacobian matrix J^t always refers to (4.5), the linearized deformation after a finite time t , either for a continuous time flow, or a discrete time mapping.

Question 4.1. Alice is confused

Q What's the difference between the *stability matrix* A and the *Jacobian matrix* J^t ?

A The velocity gradients matrix A is the *instantaneous* shear rate of a neighborhood of a point x . Dimensionally it is $(1/\text{time})$. The Jacobian matrix J^t is a dimensionless matrix of ratios of distances across the neighborhood after a *finite time* t , divided by initial distances.

Stability matrix A is a matrix of spatial derivatives of velocities, i.e., a velocity gradients matrix. J^t is obtained by a finite time integration over A .

Question 4.2. Alice asks

Q What's this $\text{tr } A = \partial_i v_i$, trace equals divergence relation described in sect. 4.7? Looks like magic to me.

A I like Mathemaniac step-by-step derivation of $\ln \det = \text{tr } \ln$ relations as flow divergences relations, see  *The real meaning of trace of matrix | Lie groups, algebras, brackets #5*

Question 4.3. Alice wants to know

Q So, computing eigenvalues and eigenvectors seems like a good thing. But how do you really do it?

A Any text on numerics of matrices discusses how this is done; the keywords are ‘Gram-Schmidt’, and for high-dimensional flows ‘Krylov subspace’ and ‘Arnoldi iteration’. Conceptually (but not for numerical purposes) we like the economical description of neighborhoods of equilibria and periodic orbits afforded by projection operators. While usually not phrased in language of projection operators, the requisite linear algebra is standard. This is a sidetrack that you will likely find confusing at the first go, so it is relegated to appendix A4.

Question 4.4. Alice does not like our Jacobian matrix

Q I do not like our definition of the Jacobian matrix in terms of the time-ordered exponential (4.19). Depending on the signs of multipliers, the left hand side of (4.28) can be either positive or negative. But the right hand side is an exponential of a real number, and that can only be positive. What gives?

A As we shall see much later on in this text, in discussion of topological indices arising in semiclassical quantization, this is not at all a dumb question.

References

- [1] G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists: A Comprehensive Guide*, 6th ed. (Academic, New York, 2005).
- [2] V. I. Arnol'd, *Ordinary Differential Equations* (Springer, New York, 1992).
- [3] I. C. Christov, R. M. Lueptow, and J. M. Ottino, “Stretching and folding versus cutting and shuffling: An illustrated perspective on mixing and deformations of continua”, *Amer. J. Phys.* **79**, 359–367 (2011).
- [4] T. F. Coleman and C. Van Loan, *Handbook for Matrix Computations* (SIAM, Philadelphia, 1988).
- [5] P. Glendinning, *Stability, Instability and Chaos: An Introduction to the Theory of Nonlinear Differential Equations* (Cambridge Univ. Press, Cambridge UK, 1994).
- [6] I. Goldhirsch, P. L. Sulem, and S. A. Orszag, “Stability and Lyapunov stability of dynamical systems: A differential approach and a numerical method”, *Physica D* **27**, 311–337 (1987).

- [7] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. (J. Hopkins Univ. Press, Baltimore, MD, 2013).
- [8] M. Gurtin, *An Introduction to Continuum Mechanics* (Academic, New York, 1981).
- [9] M. W. Hirsch and S. Smale, *Differential Equations, Dynamical Systems, and Linear Algebra* (Academic, San Diego, 1974).
- [10] P. Holmes and E. T. Shea-Brown, “Stability”, *Scholarpedia* **1**, 1838 (2006).
- [11] E. M. Izhikevich, “Equilibrium”, *Scholarpedia* **2**, 2014 (2007).
- [12] C. G. J. Jacobi, “De functionibus alternantibus earumque divisione per productum e differentiis elementorum conflatum”, *J. Reine Angew. Math. (Crelle)* **22**, 439–452 (1841).
- [13] E. N. Lorenz, “A study of the predictability of a 28-variable atmospheric model”, *Tellus* **17**, 321–333 (1965).
- [14] T. Manos, C. Skokos, and C. Antonopoulos, “Probing the local dynamics of periodic orbits by the generalized alignment index (GALI) method”, *Int. J. Bifur. Chaos* **22**, 1250218 (2012).
- [15] J. D. Meiss, *Differential Dynamical Systems* (SIAM, Philadelphia, 2007).
- [16] C. Meyer, *Matrix Analysis and Applied Linear Algebra* (SIAM, Philadelphia, 2000).
- [17] L. Perko, *Differential Equations and Dynamical Systems* (Springer, New York, 1991).
- [18] C. Skokos, “Alignment indices: a new, simple method for determining the ordered or chaotic nature of orbits”, *J. Phys. A* **34**, 10029–10043 (2001).
- [19] C. Skokos, “The Lyapunov characteristic exponents and their computation”, in *Dynamics of Small Solar System Bodies and Exoplanets*, edited by J. J. Souchay and R. Dvorak (Springer, New York, 2010), pp. 63–135.
- [20] M. Stone and P. Goldbart, *Mathematics for Physics: A Guided Tour for Graduate Students* (Cambridge Univ. Press, Cambridge UK, 2009).
- [21] M. Tabor, *Chaos and Integrability in Nonlinear Dynamics: An Introduction* (Wiley, New York, 1989).
- [22] L. N. Trefethen and D. Bau, *Numerical Linear Algebra* (SIAM, 1997).
- [23] C. A. Truesdell, *A First Course in Rational Continuum Mechanics. General Concepts*, Vol. 1 (Academic, New York, 1977).
- [24] D. S. Watkins, *The Matrix Eigenvalue Problem: GR and Krylov Subspace Methods* (SIAM, Philadelphia, 2007).
- [25] D. S. Watkins, *Fundamentals of Matrix Computations*, 3rd ed. (Wiley, New York, 2010).

4.8 Examples

10. Try to leave out the part that readers tend to skip.

— Elmore Leonard's Ten Rules of Writing.

The reader is urged to study the examples collected here. If you want to return back to the main text, click on [click to return] pointer on the margin.

Example 4.1. Rössler and Lorenz flows, linearized. (Continued from example 3.4) For the Rössler (2.30) and Lorenz (2.25) flows, the stability matrices are respectively

$$A_{Ross} = \begin{pmatrix} 0 & -1 & -1 \\ 1 & a & 0 \\ z & 0 & x - c \end{pmatrix}, \quad A_{Lor} = \begin{pmatrix} -\sigma & \sigma & 0 \\ \rho - z & -1 & -x \\ y & x & -b \end{pmatrix}. \quad (4.31)$$

(continued in example 4.5)

[return to p. 87](#)

Example 4.2. Jacobian matrix eigenvalues, diagonalizable case. If we happen to be so lucky that $A = A_D$ is a diagonal matrix with eigenvalues $(\lambda^{(1)}, \lambda^{(2)}, \dots, \lambda^{(d)})$, the exponential is simply

$$J^t = e^{tA_D} = \begin{pmatrix} e^{t\lambda^{(1)}} & \dots & 0 \\ & \ddots & \\ 0 & \dots & e^{t\lambda^{(d)}} \end{pmatrix}. \quad (4.32)$$

Next, suppose that A is diagonalizable and that U is a nonsingular matrix that brings it to a diagonal form $A_D = U^{-1}AU$. Then J can also be brought to a diagonal form (insert factors $\mathbf{1} = UU^{-1}$ between the terms of the product (4.15)):

$$J^t = e^{tA} = Ue^{tA_D}U^{-1}. \quad (4.33)$$

[exercise 4.2](#)

The action of both A and J is very simple; the axes of an orthogonal coordinate system where A is diagonal are also the eigen-directions of J^t , and under the flow the neighborhood is deformed by a multiplication by an eigenvalue factor for each coordinate axis.

[return to p. 91](#)

We recapitulate the basic facts of linear algebra in appendix A4. The following 2-dimensional example serves well to highlight the most important types of linear flows:

Example 4.3. Linear stability of 2-dimensional flows. For a 2-dimensional flow the eigenvalues $\lambda^{(1)}, \lambda^{(2)}$ of A are either real, leading to a linear motion along their eigenvectors, $x_j(t) = x_j(0) \exp(t\lambda^{(j)})$, or form a complex conjugate pair $\lambda^{(1)} = \mu + i\omega, \lambda^{(2)} = \mu - i\omega$, leading to a circular or spiral motion in the $[x_1, x_2]$ plane.

These two possibilities are refined further into sub-cases depending on the signs of the real part. In the case of real $\lambda^{(1)} > 0, \lambda^{(2)} < 0$, x_1 grows exponentially with time, and x_2 contracts exponentially. This behavior, called a *saddle*, is sketched in figure 4.3, as are the remaining possibilities: in/out nodes, inward/outward spirals, and the center. The magnitude of out-spiral $|x(t)|$ diverges exponentially when $\mu > 0$, and in-spiral contracts into $(0, 0)$ when $\mu < 0$. The phase velocity ω controls its oscillations.

[example A4.2](#)

Figure 4.3: Trajectories in linearized neighborhoods of several 2-dimensional equilibria: saddle (hyperbolic), in node (attracting), center (elliptic), in spiral.

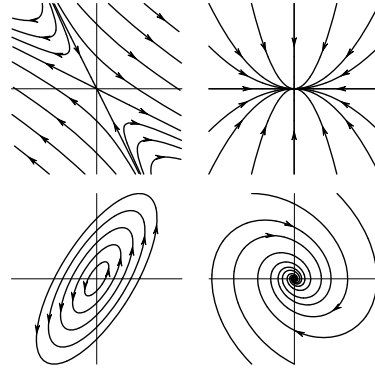
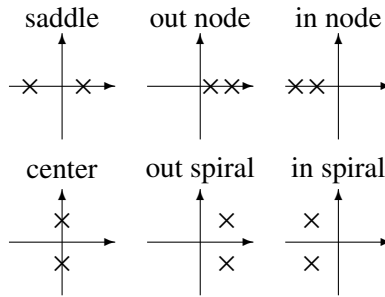


Figure 4.4: Qualitatively distinct types of stability exponents $\{\lambda^{(1)}, \lambda^{(2)}\}$, i.e., eigenvalues of the $[2 \times 2]$ stability matrix A .



If eigenvalues $\lambda^{(1)} = \lambda^{(2)} = \lambda$ are degenerate, the matrix might have two linearly independent eigenvectors, or only one eigenvector. We distinguish two cases: (a) A can be brought to diagonal form and (b) A can be brought to *Jordan form*, which (in dimension 2 or higher) has zeros everywhere except for the repeating eigenvalues on the diagonal and some 1's directly above it. For every such Jordan $[d_\alpha \times d_\alpha]$ block there is only one eigenvector per block.

[return to p. 91](#)

We sketch the full set of possibilities in figures 4.3 and 4.4, and we work out in detail the most important cases in appendix A4, example A4.2.

Example 4.4. Complex eigenvalues: in-out spirals. As $\mathbf{M}$ has only real entries, it will in general have either real eigenvalues, or complex conjugate pairs of eigenvalues. Also the corresponding eigenvectors can be either real or complex. All coordinates used in defining a dynamical flow are real numbers, so what is the meaning of a *complex* eigenvector?

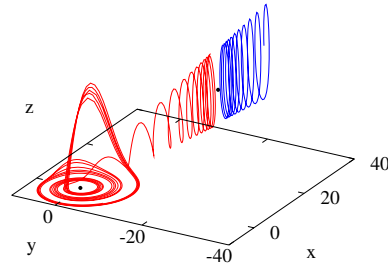
If λ_k, λ_{k+1} eigenvalues that lie within a diagonal $[2 \times 2]$ sub-block $\mathbf{M}' \subset \mathbf{M}$ form a complex conjugate pair, $\{\lambda_k, \lambda_{k+1}\} = \{\mu + i\omega, \mu - i\omega\}$, the corresponding complex eigenvectors can be replaced by their real and imaginary parts, $\{\mathbf{e}^{(k)}, \mathbf{e}^{(k+1)}\} \rightarrow \{\text{Re } \mathbf{e}^{(k)}, \text{Im } \mathbf{e}^{(k)}\}$. In this 2-dimensional real representation, $\mathbf{M}' \rightarrow A$, block A is the sum of the $\mu \mathbf{1}$ and the generator of $\text{SO}(2)$ rotations in the $\{\text{Re } \mathbf{e}^{(1)}, \text{Im } \mathbf{e}^{(1)}\}$ plane.

$$A = \begin{bmatrix} \mu & -\omega \\ \omega & \mu \end{bmatrix} = \mu \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \omega \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Trajectories of $\dot{\mathbf{x}} = A\mathbf{x}$, given by $\mathbf{x}(t) = J^t \mathbf{x}(0)$, where (omitting $\mathbf{e}^{(3)}, \mathbf{e}^{(4)}, \dots$ eigen-directions)

$$J^t = e^{tA} = e^{t\mu} \begin{bmatrix} \cos \omega t & -\sin \omega t \\ \sin \omega t & \cos \omega t \end{bmatrix}, \quad (4.34)$$

Figure 4.5: Two trajectories of the Rössler flow initiated in the neighborhood of the ‘+’ or ‘outer’ equilibrium point (2.31). (R. Paškauskas)



spiral in/out around $(x, y) = (0, 0)$, see figure 4.3, with the rotation period T and the radial expansion/contraction multiplier along the $\mathbf{e}^{(j)}$ eigen-direction per a turn of the spiral:

$$T = 2\pi/\omega, \quad \Lambda_{\text{radial}} = e^{T\mu}. \quad (4.35)$$

exercise A4.1

We learn that the typical turnover time scale in the neighborhood of the equilibrium $(x, y) = (0, 0)$ is of order $\approx T$ (and not, let us say, $1000 T$, or $10^{-2} T$).

return to p. 89

Example 4.5. Stability of equilibria of the Rössler flow. (Continued from example 4.1) The Rössler system (2.30) has two equilibrium points (2.31), the inner equilibrium (x_-, y_-, z_-) , and the outer equilibrium point (x^+, y^+, z^+) . Together with their exponents (eigenvalues of the stability matrix), the two equilibria yield quite detailed information about the flow. Figure 4.5 shows two trajectories which start in the neighborhood of the outer ‘+’ equilibrium. Trajectories to the right of the equilibrium point ‘+’ escape, and those to the left spiral toward the inner equilibrium point ‘-’, where they seem to wander chaotically for all times. The stable manifold of the outer equilibrium point thus serves as the attraction basin boundary. Consider now the numerical values for eigenvalues of the two equilibria:

exercise 4.4
exercise 2.8

$$\begin{aligned} (\mu_-^{(1)}, \mu_-^{(2)} \pm i\omega_-^{(2)}) &= (-5.686, \quad 0.0970 \pm i0.9951) \\ (\mu_+^{(1)}, \mu_+^{(2)} \pm i\omega_+^{(2)}) &= (0.1929, \quad -4.596 \times 10^{-6} \pm i5.428). \end{aligned} \quad (4.36)$$

Outer equilibrium: The $\mu_+^{(2)} \pm i\omega_+^{(2)}$ complex eigenvalue pair implies that the neighborhood of the outer equilibrium point rotates with angular period $T_+ \approx |2\pi/\omega_+^{(2)}| = 1.1575$. The multiplier by which a trajectory that starts near the ‘+’ equilibrium point contracts in the stable manifold plane is the excruciatingly slow multiplier $\Lambda_2^+ \approx \exp(\mu_+^{(2)} T_+) = 0.9999947$ per rotation. For each period, the point of the stable manifold moves away along the unstable eigen-direction by factor $\Lambda_1^+ \approx \exp(\mu_+^{(1)} T_+) = 1.2497$. Hence the slow spiraling on both sides of the ‘+’ equilibrium point.

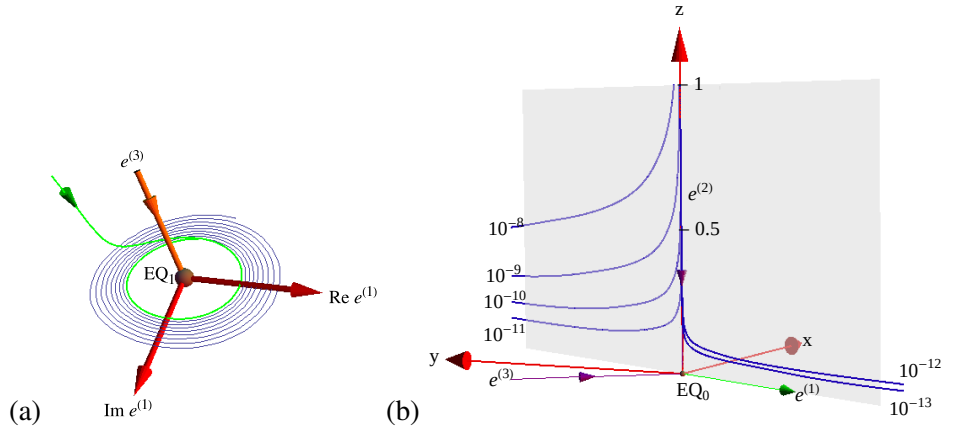
Inner equilibrium: The $\mu_-^{(2)} \pm i\omega_-^{(2)}$ complex eigenvalue pair tells us that the neighborhood of the ‘-’ equilibrium point rotates with angular period $T_- \approx |2\pi/\omega_-^{(2)}| = 6.313$, slightly faster than the harmonic oscillator estimate in (2.27). The multiplier, by which a trajectory that starts near the ‘-’ equilibrium point spirals away per one rotation, is $\Lambda_{\text{radial}} \approx \exp(\mu_-^{(2)} T_-) = 1.84$. The $\mu_-^{(1)}$ eigenvalue is essentially the z expansion correcting parameter c introduced in (2.29). For each Poincaré section return, the trajectory is contracted into the stable manifold by the amazing factor of $\Lambda_1 \approx \exp(\mu_-^{(1)} T_-) = 10^{-15.6}$.

Suppose you start with a 1 mm interval pointing in the Λ_1 eigen-direction. After one Poincaré return, the interval is of the order of 10^{-4} fermi, the furthest we will get into subnuclear structure in this book. Of course, from the mathematical point of view, the flow is reversible, and the return map is invertible. (continued in example 14.3)

return to p. 92

(R. Paškauskas)

Figure 4.6: (a) A perspective view of the linearized Lorenz flow near EQ_1 equilibrium, see figure 3.5 (a). The unstable eigenplane of EQ_1 is spanned by $\{\text{Re } \mathbf{e}^{(1)}, \text{Im } \mathbf{e}^{(1)}\}$; the stable subspace by the stable eigenvector $\mathbf{e}^{(3)}$. (b) Lorenz flow near the EQ_0 equilibrium: unstable eigenvector $\mathbf{e}^{(1)}$, stable eigenvectors $\mathbf{e}^{(2)}, \mathbf{e}^{(3)}$. Trajectories, initiated at distances $10^{-8} \dots 10^{-12}, 10^{-13}$ away from the z -axis, exit finite distance from EQ_0 along the $(\mathbf{e}^{(1)}, \mathbf{e}^{(2)})$ eigenvectors plane. Due to the strong $\lambda^{(1)}$ expansion, the EQ_0 equilibrium is, for all practical purposes, unreachable, and the $EQ_0 \rightarrow EQ_1$ heteroclinic connection is never observed in simulations such as figure 2.5. (E. Siminos; continued in figure 14.14.)



Example 4.6. Stability of Lorenz flow equilibria. (Continued from example 4.1) A glance at figure 3.5 suggests that the flow is organized by its 3 equilibria, so let us have a closer look at their stable/unstable manifolds.

The EQ_0 equilibrium stability matrix (4.31) evaluated at $x_{EQ_0} = (0, 0, 0)$ is block-diagonal. The z -axis is an eigenvector with a contracting eigenvalue $\lambda^{(2)} = -b$. From (4.42) it follows that all $[x, y]$ areas shrink at the rate $-(\sigma + 1)$. Indeed, the $[x, y]$ submatrix

remark 11.8

$$A^- = \begin{pmatrix} -\sigma & \sigma \\ \rho & -1 \end{pmatrix} \quad (4.37)$$

has a real expanding/contracting eigenvalue pair, $\lambda^{(1,3)} = -(\sigma + 1)/2 \pm \sqrt{(\sigma - 1)^2/4 + \rho\sigma}$, with the right eigenvectors $\mathbf{e}^{(1)}, \mathbf{e}^{(3)}$ in the $[x, y]$ plane, given by (either) column of the projection operator

$$P_i = \frac{A^- - \lambda^{(j)} \mathbf{1}}{\lambda^{(i)} - \lambda^{(j)}} = \frac{1}{\lambda^{(i)} - \lambda^{(j)}} \begin{pmatrix} -\sigma - \lambda^{(j)} & \sigma \\ \rho & -1 - \lambda^{(j)} \end{pmatrix}, \quad i \neq j \in \{1, 3\}. \quad (4.38)$$

$EQ_{1,2}$ equilibria have no symmetry, so their eigenvalues are given by the roots of a cubic equation, the secular determinant $\det(A - \lambda \mathbf{1}) = 0$:

$$\lambda^3 + \lambda^2(\sigma + b + 1) + \lambda b(\sigma + \rho) + 2\sigma b(\rho - 1) = 0. \quad (4.39)$$

For $\rho > 24.74$, $EQ_{1,2}$ have one stable real eigenvalue and one unstable complex conjugate pair, leading to a spiral-out instability and the strange attractor depicted in figure 2.5.

All numerical plots of the Lorenz flow are carried out here with the Lorenz parameters set to $\sigma = 10$, $b = 8/3$, $\rho = 28$. We note the corresponding stability exponents for future reference,

$$\begin{aligned} EQ_0 : (\lambda^{(1)}, \lambda^{(2)}, \lambda^{(3)}) &= (11.83, -2.666, -22.83) \\ EQ_1 : (\mu^{(1)} \pm i\omega^{(1)}, \lambda^{(3)}) &= (0.094 \pm i 10.19, -13.85). \end{aligned} \quad (4.40)$$

We also note the rotation period $T_{EQ_1} = 2\pi/\omega^{(1)}$ about EQ_1 and the associated expansion/contraction multipliers $\Lambda^{(i)} = \exp(\mu^{(j)} T_{EQ_1})$ per spiral-out turn:

$$T_{EQ_1} = 0.6163, \quad (\Lambda^{(1)}, \Lambda^{(3)}) = (1.060, 1.957 \times 10^{-4}). \quad (4.41)$$

We learn that the typical turnover time scale in this problem is of the order $T \approx T_{EQ_1} \approx 1$ (and not, let us say, 1000, or 10^{-2}). Combined with the contraction rate (4.42), this tells

us that the Lorenz flow strongly contracts state space volumes, by factor of $\approx 10^{-4}$ per mean turnover time.

In the EQ_1 neighborhood, the unstable manifold trajectories slowly spiral out, with a very small radial per-turn expansion multiplier $\Lambda^{(1)} \simeq 1.06$ and a very strong contraction multiplier $\Lambda^{(3)} \simeq 10^{-4}$ onto the unstable manifold, figure 4.6 (a). This contraction confines, for all practical purposes, the Lorenz attractor to a 2-dimensional surface, which is evident in figure 3.5.

In the $x_{EQ_0} = (0, 0, 0)$ equilibrium neighborhood, the extremely strong $\lambda^{(3)} \simeq -23$ contraction along the $\mathbf{e}^{(3)}$ direction confines the hyperbolic dynamics near EQ_0 to the plane spanned by the unstable eigenvector $\mathbf{e}^{(1)}$, with $\lambda^{(1)} \simeq 12$, and the slowest contraction rate eigenvector $\mathbf{e}^{(2)}$ along the z -axis, with $\lambda^{(2)} \simeq -3$. In this plane, the strong expansion along $\mathbf{e}^{(1)}$ overwhelms the slow $\lambda^{(2)} \simeq -3$ contraction down the z -axis, making it extremely unlikely for a random trajectory to approach EQ_0 , figure 4.6 (b). Thus, linearization describes analytically both the singular dip in the Poincaré sections of figure 3.5 and the empirical scarcity of trajectories close to EQ_0 . (continued in example 4.8)

[return to p. 92](#)

(E. Siminos and J. Halcrow)

Example 4.7. Lorenz flow: A global portrait. (Continued from example 4.6) As the EQ_1 unstable manifold spirals out, the strip that starts out in the section above EQ_1 in figure 3.5 cuts across the z -axis invariant subspace. This strip necessarily contains a heteroclinic orbit that hits the z -axis head on, and in infinite time (but exponentially fast) descends all the way to EQ_0 .

How? Since the dynamics is linear (see figure 4.6 (a)) in the neighborhood of EQ_0 , there is no need to integrate numerically the final segment of the heteroclinic connection. It is sufficient to bring a trajectory a small distance away from EQ_0 , continue analytically to a small distance beyond EQ_0 and then resume the numerical integration.

What happens next? Trajectories to the left of the z -axis shoot off along the $\mathbf{e}^{(1)}$ direction, and those to the right along $-\mathbf{e}^{(1)}$. Given that $xy > 0$ along the $\mathbf{e}^{(1)}$ direction, the nonlinear term in the $\dot{z}$ equation (2.25) bends both branches of the EQ_0 unstable manifold $W^u(EQ_0)$ upwards. Then ... - never mind. We postpone completion of this narrative to example 11.8, where the discrete symmetry of Lorenz flow will help us streamline the analysis. As we shall show, what we already know about the 3 equilibria and their stable/unstable manifolds suffices to completely pin down the topology of Lorenz flow. (continued in example 11.8)

[return to p. 92](#)

(E. Siminos and J. Halcrow)

Example 4.8. Lorenz flow state space contraction. (Continued from example 4.6) It follows from (4.31) and (4.29) that Lorenz flow is volume contracting,

$$\partial_i v_i = \sum_{i=1}^3 \lambda^{(i)}(x, t) = -\sigma - b - 1, \quad (4.42)$$

at a constant, coordinate- and ρ -independent rate, set by Lorenz to $\partial_i v_i = -13.66$. For periodic orbits and long time averages, there is no contraction/expansion along the flow, $\lambda^{(ll)} = 0$, and the sum of $\lambda^{(i)}$ is constant by (4.42). Thus, we compute only one independent exponent $\lambda^{(i)}$. (continued in example 11.8)

[return to p. 95](#)

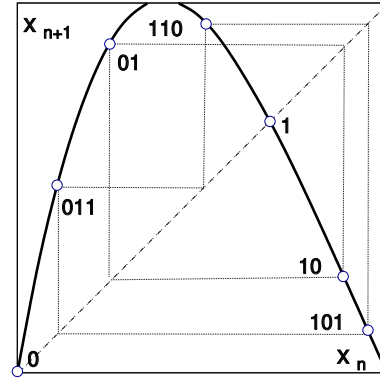


Figure 4.7: A unimodal map, together with fixed points 0, 1, 2-cycle 01 and 3-cycle 011.

Example 4.9. Stability of a 1-dimensional map. Consider the orbit $\{\dots, x_{-1}, x_0, x_1, x_2, \dots\}$ of a 1-dimensional map $x_{n+1} = f(x_n)$. When studying linear stability (and higher derivatives) of the map, it is often convenient to use a local coordinate system z_a centered on the orbit point x_a , together with a notation for the map, its derivative, and, by the chain rule, the derivative of the k th iterate f^k evaluated at the point x_a ,

$$\begin{aligned} x &= x_a + z_a, & f_a(z_a) &= f(x_a + z_a) \\ f'_a &= f'(x_a) \\ \Lambda(x_0, k) &= f_a^{k'} = f'_{a+k-1} \cdots f'_{a+1} f'_a, & k &\geq 2. \end{aligned} \quad (4.43)$$

Here a is the label of point x_a , and the label $a+1$ is shorthand for the next point b on the orbit of x_a , $x_b = x_{a+1} = f(x_a)$. For example, a period-3 periodic point in figure 4.7 might have label $a = 011$, and by $x_{110} = f(x_{011})$ the next point label is $b = 110$.

[return to p. 93](#)

Example 4.10. Hénon map Jacobian matrix. For the Hénon map (3.18) the Jacobian matrix for the n th iterate of the map is

$$M^n(x_0) = \prod_{m=n}^1 \begin{bmatrix} -2ax_m & b \\ 1 & 0 \end{bmatrix}, \quad x_m = f_1^m(x_0, y_0). \quad (4.44)$$

The determinant of the Hénon one time-step Jacobian matrix (4.44) is constant,

$$\det M = \Lambda_1 \Lambda_2 = -b. \quad (4.45)$$

In this case only one eigenvalue $\Lambda_1 = -b/\Lambda_2$ needs to be determined. This is not an accident; a constant Jacobian was one of desiderata that led Hénon to construct a map of this particular form.

[return to p. 93](#)

4.9 Exercises

4.1. **Trace-log of a matrix.** Prove that

$$\det M = e^{\text{tr} \ln M}.$$

for an arbitrary nonsingular finite dimensional matrix M , $\det M \neq 0$.

4.2. **Stability, diagonal case.** Verify the relation (4.33)

$$J^t = e^{tA} = \mathbf{U}^{-1} e^{tA_D} \mathbf{U}, \quad A_D = \mathbf{U} \mathbf{A} \mathbf{U}^{-1}.$$

4.3. **State space volume contraction.**

- Compute the Rössler flow volume contraction rate at the equilibria.
- Study numerically the instantaneous $\partial_i v_i$ along a typical trajectory on the Rössler attractor; color-code the points on the trajectory by the sign (and perhaps the magnitude) of $\partial_i v_i$. If you see regions of local expansion, explain them.
- (optional) Color-code the points on the trajectory by the sign (and perhaps the magnitude) of $\partial_i v_i - \partial_i \bar{v}_i$.
- Compute numerically the average contraction rate (4.29) along a typical trajectory on the Rössler attractor. Plot it as a function of time.
- Argue on the basis of your results that this attractor has a smaller dimension than the state space $d = 3$.
- (optional) Start some trajectories on the escape side of the outer equilibrium, and color-code the points on the trajectory. Is the flow volume contracting?

(continued in exercise 23.10)

4.4. **Topology of the Rössler flow.** (continuation of exercise 3.1)

- Show that equation $|\det(A - \lambda \mathbf{I})| = 0$ for Rössler flow in the notation of exercise 2.8 can be written as

$$\lambda^3 + \lambda^2 c (p^\mp - \epsilon) + \lambda(p^\pm/\epsilon + 1 - c^2 \epsilon p^\mp) \mp c \sqrt{D} = 0 \quad (4.46)$$

- Solve (4.46) for eigenvalues $\lambda^\pm$ for each equilibrium as an expansion in powers of ϵ . Derive

$$\begin{aligned} \lambda_1^- &= -c + \epsilon c/(c^2 + 1) + o(\epsilon) \\ \lambda_2^- &= \epsilon c^3/[2(c^2 + 1)] + o(\epsilon^2) \\ \theta_2^- &= 1 + \epsilon/[2(c^2 + 1)] + o(\epsilon) \\ \lambda_1^+ &= c\epsilon(1 - \epsilon) + o(\epsilon^3) \\ \lambda_2^+ &= -\epsilon^5 c^2/2 + o(\epsilon^6) \\ \theta_2^+ &= \sqrt{1 + 1/\epsilon} (1 + o(\epsilon)) \end{aligned} \quad (4.47)$$

Compare with exact eigenvalues. What are dynamical implications of the extravagant value of λ_1^- ? (continued as exercise 7.1)

(R. Paškauskas)

- 4.5. **Time-ordered exponentials.** Given a time dependent matrix $A(t)$ check that the time-ordered exponential

$$J(t) = \mathbf{T}e^{\int_0^t d\tau A(\tau)}$$

may be written as

$$J(t) = \sum_{m=0}^{\infty} \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{m-1}} dt_m A(t_1) \cdots A(t_m)$$

and verify, by using this representation, that $J(t)$ satisfies the equation

$$\dot{J}(t) = A(t)J(t),$$

with the initial condition $J(0) = 1$.

- 4.6. **A contracting baker's map.** Consider a contracting (or 'dissipative') baker's map, acting on a unit square $[0, 1]^2 = [0, 1] \times [0, 1]$, defined by

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} x_n/3 \\ 2y_n \end{pmatrix} \quad y_n \leq 1/2$$

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} x_n/3 + 1/2 \\ 2y_n - 1 \end{pmatrix} \quad y_n > 1/2.$$

This map shrinks strips by a factor of $1/3$ in the x -direction, and then it stretches (and folds) them by a factor of 2 in the y -direction.

By how much does the state space volume contract for one iteration of the map? (continued in exercise 14.3)

Chapter 5

Cycle stability

We owe it to a book to withhold judgment until we reach page 100.

—Henrietta McNutt, George Johnson's seventh-grade English teacher

THOPOLOGICAL FEATURES of a dynamical system –singularities, periodic orbits, and the ways in which the orbits intertwine– are invariant under a general continuous change of coordinates. Equilibria and periodic orbits are *flow-invariant* sets, in the sense that the flow only shifts points along a periodic orbit, but the periodic orbit as the set of periodic points remains unchanged in time. Surprisingly, there also exist quantities that depend on the notion of metric distance between points, but nevertheless do not change value under a smooth change of coordinates. Local quantities, such as the eigenvalues of equilibria and periodic orbits and global quantities, such as Lyapunov exponents, metric entropy, and fractal dimensions, are examples of dynamical system properties that are independent of coordinate choice.

We now turn to the first, local class of such invariants, linear stability of equilibria and periodic orbits of flows and maps. This will give us metric information about local dynamics, as well as a key concept regarding the *neighborhood* around point x . The size of a neighborhood is primarily determined by the number of expanding directions and their rates of expansion; contracting directions play only a secondary role (see sect. 5.6).

If you already know that eigenvalues of periodic orbits are invariants of a flow, skip this chapter.



fast track:
chapter 7, p. 136

As noted on page 43, a trajectory can be stationary, periodic or aperiodic. For chaotic systems almost all trajectories are aperiodic–nevertheless, equilibria and

periodic orbits turn out to be the key to unraveling chaotic dynamics. Here we note a few of the properties that make them so precious to a theorist.

5.1 Equilibria

At the still point, there the dance is.

—T. S. Eliot, *Four Quartets - Burnt Norton* [00:15:30]



For a start, consider the case where x_q is an equilibrium point (2.11). Expanding around the equilibrium point x_q , using the fact that the stability matrix $A = A(x_q)$ in (4.2) is constant, and integrating, $f^t(x) = x_q + e^{At}(x - x_q) + \dots$, we verify that the simple formula (4.15) applies also to the Jacobian matrix of an equilibrium point, 

$$J_q^t = e^{A_q t}, \quad J_q^t = J^t(x_q), \quad A_q = A(x_q). \quad (5.1)$$

When an equilibrium point is stationary, time plays no role. The eigenvalues and the eigenvectors of the stability matrix A_q evaluated at the equilibrium point x_q ,

$$A_q \mathbf{e}^{(j)} = \lambda_q^{(j)} \mathbf{e}^{(j)}, \quad (5.2)$$

describe the linearized neighborhood of the equilibrium point, with stability exponents $\lambda_q^{(j)} = \mu_q^{(j)} \pm i\omega_q^{(j)}$ independent of any particular coordinate choice. Assume that these eigenvalues are non-degenerate, $\lambda^{(j)} \neq \lambda^{(k)}$ for any pair of eigenvalues.

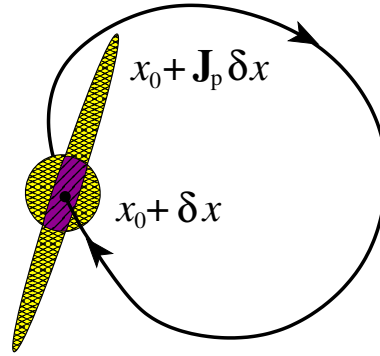
- If all $\mu^{(j)} < 0$, then the equilibrium is *stable*, or a *sink*. For $\omega^{(j)} = 0$, it is an *in-node*; for $\omega^{(j)} \neq 0$, it is an *in-spiral*.
- If some $\mu^{(j)} < 0$, and other $\mu^{(j)} > 0$, the equilibrium is *hyperbolic*, or a *saddle*.
- If all $\mu^{(j)} > 0$, then the equilibrium is *repelling*, or a *source*. For $\omega^{(j)} = 0$, it is an *out-node*; for $\omega^{(j)} \neq 0$, it is an *out-spiral*.
- If some $\mu^{(j)} = 0$, think again (you have a symmetry or a bifurcation).

For 2-dimensional flows, these types of equilibrium stabilities are illustrated in figures 4.4 and 4.3. The stability matrix eigenvectors (5.2) are the eigenvectors of the Jacobian matrix as well, $J_q^t \mathbf{e}^{(j)} = \exp(t\lambda_q^{(j)}) \mathbf{e}^{(j)}$.

5.2 Periodic orbits

An obvious virtue of periodic orbits is that they are *topological* invariants. A fixed point remains fixed for any choice of coordinates, and similarly a periodic orbit remains periodic in any representation of the dynamics. Any re-parametrization

Figure 5.1: For a prime cycle p , Floquet matrix J_p returns an infinitesimal spherical neighborhood of $x_0 \in \mathcal{M}_p$ stretched into an ellipsoid, with overlap ratio along the eigendirection $\mathbf{e}^{(j)}$ of $J_p(x)$ given by the Floquet multiplier $|\Lambda_j|$. These ratios are invariant, under smooth nonlinear reparametrizations of state space coordinates, and also represent an intrinsic property of cycle p .



of a dynamical system that preserves its topology has to preserve topological relations between periodic orbits, such as their relative inter-windings and knots. So the mere existence of periodic orbits suffices to partially organize the spatial layout of a non-wandering set. No less important, as we shall now show, is the fact that cycle eigenvalues are *metric* invariants; they determine the relative sizes of neighborhoods in a non-wandering set.

We start by noting that due to the multiplicative structure (4.20) of Jacobian matrices, the Jacobian matrix for the r th repeat of a prime cycle p of period T is

$$J^{rT}(x) = J^T(f^{(r-1)T}(x)) \cdots J^T(f^T(x))J^T(x) = J_p(x)^r, \quad (5.3)$$

where $J_p(x) = J^T(x)$ is the Jacobian matrix for a single traversal of the prime cycle p . When $x \in \mathcal{M}_p$, x is any point on the cycle, and $f^{rT}(x) = x$ as $f^t(x)$ returns to x for every multiple of period T . Hence, it suffices to restrict our considerations to the stability of prime cycles.



fast track:
sect. 5.3, p. 113

5.2.1 Cycle stability



The time-dependent T -periodic vector fields, such as the flow linearized around a periodic orbit, are described by Floquet theory. Hence from now on we shall refer to a Jacobian matrix evaluated on a periodic orbit p either as a $[d \times d]$ *Floquet matrix* J_p or a $[(d-1) \times (d-1)]$ *monodromy matrix* M_p , to its eigenvalues Λ_j as *Floquet multipliers* (4.7), and to

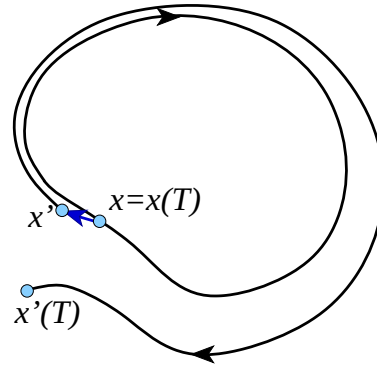
appendix A4.2.1

$$\lambda_p^{(j)} = \mu_p^{(j)} + i\omega_p^{(j)} \quad (5.4)$$

as *Floquet exponents*. In the literature they are sometimes called “characteristic” multipliers and exponents. The stretching/contraction rates per unit time are given by the real parts of Floquet exponents

$$\mu_p^{(j)} = \frac{1}{T_p} \ln |\Lambda_{p,j}|. \quad (5.5)$$

Figure 5.2: An unstable periodic orbit repels every neighboring trajectory $x'(t)$, except those on its center and stable manifolds.



The factor $1/T_p$ in the definition of the Floquet exponents is motivated by its form for the linear dynamical systems, for example (4.32). (Parenthetically, a Floquet exponent is not a Lyapunov exponent (6.12) evaluated on one period of prime cycle p ; the relation is subtler than that, read chapter 6). When Λ_j is real, we do care about $\sigma^{(j)} = \Lambda_j/|\Lambda_j| \in \{+1, -1\}$, the sign of the j th Floquet multiplier. If $\sigma^{(j)} = -1$ and $|\Lambda_j| \neq 1$, the corresponding eigen-direction is said to be *inverse hyperbolic*. Keeping track of this case-by-case enumeration is an unnecessary nuisance, so most of our formulas will be stated in terms of the Floquet multipliers Λ_j rather than in the terms of the multiplier signs $\sigma^{(j)}$, exponents $\mu^{(j)}$ and phases $\omega^{(j)}$.

section 8.4

In dynamics the expanding directions, $|\Lambda_e| > 1$, need to be addressed first, while the contracting directions $|\Lambda_c| < 1$ tend to resolve themselves. Hence, we always order multipliers Λ_k in order of decreasing magnitude $|\Lambda_1| \geq |\Lambda_2| \geq \dots \geq |\Lambda_d|$. Since $|\Lambda_j| = e^{t\mu^{(j)}}$, this is the same as ordering by $\mu^{(1)} \geq \mu^{(2)} \geq \dots \geq \mu^{(d)}$. We sort the Floquet multipliers $\{\Lambda_{p,1}, \Lambda_{p,2}, \dots, \Lambda_{p,d}\}$ of the Floquet matrix evaluated on the p -cycle into three sets $\{e, m, c\}$

$$\begin{aligned}
 \text{expanding:} \quad \{\Lambda\}_e &= \{\Lambda_{p,j} : |\Lambda_{p,j}| > 1\} \\
 &\quad \{\lambda\}_e = \{\lambda_p^{(j)} : \mu_p^{(j)} > 0\} \\
 \text{marginal:} \quad \{\Lambda\}_m &= \{\Lambda_{p,j} : |\Lambda_{p,j}| = 1\} \\
 &\quad \{\lambda\}_m = \{\lambda_p^{(j)} : \mu_p^{(j)} = 0\} \\
 \text{contracting:} \quad \{\Lambda\}_c &= \{\Lambda_{p,j} : |\Lambda_{p,j}| < 1\} \\
 &\quad \{\lambda\}_c = \{\lambda_p^{(j)} : \mu_p^{(j)} < 0\}.
 \end{aligned} \tag{5.6}$$

In what follows, the volume of an expanding manifold plays an important role. The product of the *expanding* Floquet multipliers is denoted by Λ_p (no j th eigenvalue index),

$$\Lambda_p = \prod_e \Lambda_{p,e}. \tag{5.7}$$

As J_p is a real matrix, complex eigenvalues always come in complex conjugate pairs, $\Lambda_{p,i+1} = \Lambda_{p,i}^*$, so the product (5.7) is always real.

A periodic orbit of a continuous-time flow, map, or fixed point on a map is

p. 110

- *stable, attracting, a sink or a limit cycle* if all $|\Lambda_j| < 1$ (real parts of all of its Floquet exponents, other than the vanishing longitudinal exponent for perturbations tangent to the cycle, see sect. 5.3.1, are strictly negative, $0 > \mu^{(1)} \geq \mu^{(j)}$).
- *hyperbolic or saddle*, unstable to perturbations outside its stable manifold if some $|\Lambda_j| > 1$, and other $|\Lambda_j| < 1$ (a set of $\mu^{(j)} \geq \mu_{min} > 0$ is strictly positive, and the rest is strictly negative).
- *elliptic, neutral or marginal* if all $|\Lambda_j| = 1$ ($\mu^{(j)} = 0$).
- *partially hyperbolic*, if $\mu^{(j)} = 0$ for a subset of exponents (other than the longitudinal one).
- *repelling or source*, unstable to any perturbation if *all* $|\Lambda_j| > 1$ (all Floquet exponents, other than the vanishing longitudinal exponent, are strictly positive, $\mu^{(j)} \geq \mu^{(d)} > 0$).

The region of system parameter values for which a periodic orbit p is stable is called the *stability window* of p . The set of initial points that are asymptotically attracted to M_p as $t \rightarrow +\infty$ (for a fixed set of system parameter values) is called the *basin of attraction* of limit cycle p . Repelling and hyperbolic cycles are unstable to generic perturbations, and thus said to be *unstable*, see figure 5.2.

section 8.5

If *all* Floquet exponents (other than the vanishing longitudinal exponent) of *all* periodic orbits of a flow are strictly bounded away from zero, the flow is said to be *hyperbolic*. Otherwise the flow is said to be *nonhyperbolic*. A confined smooth flow or map is generically nonhyperbolic, with partial ellipticity or marginality expected only in the presence of continuous symmetries, or for bifurcation parameter values. As we shall see in chapter 12, in the presence of continuous symmetries, equilibria and periodic orbits are not likely solutions, and their role is played by higher-dimensional tori, relative equilibria and relative periodic orbits. For Hamiltonian flows the symplectic $\text{Sp}(d)$ symmetry (Liouville phase-space volume conservation, Poincaré invariants) leads to a proliferation of elliptic and partially hyperbolic tori.

section 8.6
question 5.1



example 5.1
p. 122



example 5.2
p. 122

5.3 Floquet multipliers are invariant



As already noted in (5.1), if the stability matrix $A(x)$ is computed on an equilibrium point q ,

$$A_q = A(x_q), \quad (5.8)$$

its eigenvalues $\lambda_q^{(k)}$ are flow- and coordinate transformations invariant, so we label them by q , omit (x_q) .

The 1-dimensional map Floquet multiplier (5.23) is a product of derivatives over all points around the cycle, and is therefore independent of which periodic point is chosen as the initial one. In higher dimensions the form of the Floquet matrix $J_p(x_0)$ in (5.3) does depend on the choice of coordinates and the initial point $x_0 \in \mathcal{M}_p$. Nevertheless, as we shall now show, the cycle *Floquet multipliers* are intrinsic property of a cycle in any dimension. Consider the i th eigenvalue, eigenvector pair $(\Lambda_j, \mathbf{e}^{(j)})$ computed from J_p evaluated at a periodic point x ,

$$J_p(x) \mathbf{e}^{(j)}(x) = \Lambda_j \mathbf{e}^{(j)}(x), \quad x \in \mathcal{M}_p. \quad (5.9)$$

Consider another point on the cycle at time t later, $x' = f^t(x)$ whose Floquet matrix is $J_p(x')$. By the semigroup property (4.20), $J^{T+t} = J^{t+T}$, the Jacobian matrix at x' can be written either as

$$J^{T+t}(x) = J^T(x') J^t(x) = J_p(x') J^t(x),$$

or $J^t(x) J_p(x)$. Multiplying (5.9) by $J^t(x)$, we find that the Floquet matrix evaluated at x' has the same Floquet multiplier,

$$J_p(x') \mathbf{e}^{(j)}(x') = \Lambda_j \mathbf{e}^{(j)}(x'), \quad \mathbf{e}^{(j)}(x') = J^t(x) \mathbf{e}^{(j)}(x), \quad (5.10)$$

but with the eigenvector $\mathbf{e}^{(j)}$ transported along the flow $x \rightarrow x'$ to $\mathbf{e}^{(j)}(x') = J^t(x) \mathbf{e}^{(j)}(x)$. Hence, in the spirit of the Floquet theory (appendix A4.2.1) one can define time-periodic eigenvectors (in a co-moving ‘Lagrangian frame’)

$$\mathbf{e}^{(j)}(t) = e^{-\lambda^{(j)}t} J^t(x) \mathbf{e}^{(j)}(0), \quad \mathbf{e}^{(j)}(t) = \mathbf{e}^{(j)}(x(t)), \quad x(t) \in \mathcal{M}_p. \quad (5.11)$$

J_p evaluated anywhere along the cycle has the same set of Floquet multipliers $\{\Lambda_1, \Lambda_2, \dots, 1, \dots, \Lambda_{d-1}\}$. Quantities such as $\text{tr } J_p(x)$, $\det J_p(x)$ depend only on the eigenvalues of $J_p(x)$ and not on the starting point x . Thus, in expressions such as $\det(\mathbf{1} - J_p^r(x))$, we may omit reference to x ,

$$\det(\mathbf{1} - J_p^r) = \det(\mathbf{1} - J_p^r(x)) \quad \text{for any } x \in \mathcal{M}_p. \quad (5.12)$$

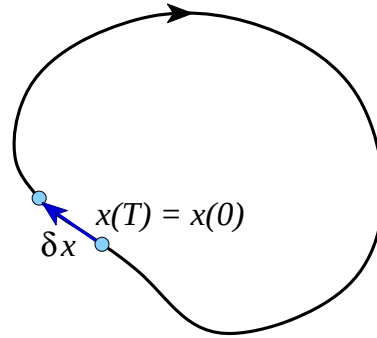
We postpone the proof that the cycle Floquet multipliers are smooth conjugacy invariants of the flow to sect. 5.4; the time-forward map (5.10) is a special case of this general property of smooth manifolds and their tangent spaces.

5.3.1 Marginal eigenvalues

The presence of marginal eigenvalues signals either continuous symmetry of a flow (which one should immediately exploit to simplify the problem), or nonhyperbolicity of a flow (a source of much pain, hard to avoid). In that case (typical of parameter values for which bifurcations occur), one has to go beyond linear stability and Jordan type subspaces (see example 4.3), as well as examine sub-exponential growth rates, such as t^α . For flow-invariant solutions such as periodic orbits, the time evolution is itself a continuous symmetry; hence a periodic orbit of a flow always has a *marginal Floquet multiplier*, as we now show.



Figure 5.3: Any two points along a periodic orbit p are mapped into themselves after one cycle period T ; hence, a longitudinal deviation vector $\delta x = v(x_0)\delta t$ is mapped into itself by the cycle Jacobian matrix J_p .



The Jacobian matrix $J^t(x)$ transports the velocity field $v(x)$ by (4.9), $v(x(t)) = J^t(x_0)v(x_0)$. In general the velocity at point $x(t)$ does not point in the same direction as the velocity at point x_0 , so this is not an eigenvalue condition for J^t . The Jacobian matrix computed for an arbitrary segment of an arbitrary trajectory has no invariant meaning. However, if the orbit is periodic, $x(T_p) = x(0)$, after a complete period

$$J_p(x)v(x) = v(x), \quad x \in \mathcal{M}_p. \quad (5.13)$$

Two successive points on the cycle, with an initial distance $\delta x = x'(0) - x(0)$ apart, are separated by exactly the same distance after a completed period $\delta x(T) = \delta x$, see figure 5.3. Thus, the velocity field v at any point along the cycle is an eigenvector $\mathbf{e}^{(\parallel)}(x) = v(x)$ of the Jacobian matrix J_p with the unit Floquet multiplier, zero Floquet exponent

$$\Lambda_{\parallel} = 1, \quad \lambda^{(\parallel)} = 0. \quad (5.14)$$

The continuous invariance that gives rise to this marginal Floquet multiplier is the invariance of a cycle (the set $\mathcal{M}_p$) under a time translation of its points along the cycle. As we shall see in sect. 5.5, this marginal stability direction can be eliminated by cutting the cycle by a Poincaré section and replacing the continuous flow Floquet matrix by the Floquet matrix of the return map.

If the flow is governed by a time-independent Hamiltonian, the energy is conserved, and that leads to an additional marginal Floquet multiplier (we shall show in sect. 8.5 that due to the symplectic invariance (8.26) real eigenvalues come in pairs). Further marginal eigenvalues arise in presence of continuous symmetries, as discussed in chapter 12.

5.4 Floquet multipliers are metric invariants

In sect. 5.3 we established that for a given flow, the Floquet multipliers are intrinsic to a given cycle, independent of the starting point along the cycle. Now we prove a much stronger statement: cycle Floquet multipliers are *smooth conjugacy* or *metric invariants* of the flow, the same in *any* representation of the dynamical system. This statement is based on concepts in elementary differential geometry, such as those discussed in the following paragraphs.



If the same dynamics is given by a map f in x coordinates, and a map g in the $y = h(x)$ coordinates, then f and g (or any other good representation) are related by a *smooth conjugacy*. Thus, the reparameterization and coordinate transformation $g = h \circ f \circ h^{-1}$ maps nearby points of f into nearby points of g . As both f and g are arbitrary representations of the dynamical system, the explicit form of the conjugacy h is of no interest, only the properties invariant under any transformation h are of general import. Furthermore, a good representation should not mutilate the data; the mapping h must be a *smooth conjugacy* which maps nearby points of f into nearby points of g .

This smoothness guarantees that the cycles are not only topological invariants, but that their linearized neighborhoods are also metric invariants. For a fixed point $f(x) = x$ of a 1-dimensional map, this follows from the chain rule for derivatives,

$$\begin{aligned} g'(y) &= h'(f \circ h^{-1}(y)) f'(h^{-1}(y)) \frac{1}{h'(x)} \\ &= h'(x) f'(x) \frac{1}{h'(x)} = f'(x). \end{aligned} \quad (5.15)$$



In d dimensions the relationship between the maps in different coordinate representations is again $g \circ h = h \circ f$. The chain rule now relates J' , the Jacobian matrix of the map g , to the Jacobian matrix of map f :

$$J'(y)_{ij} = \Gamma(f(x))_{ik} J(x)_{kl} \Gamma(x)^{-1}_{lj}, \quad (5.16)$$

where the coordinate transformation Jacobian matrices are

$$\Gamma(x)_{ik} = \frac{\partial y_i}{\partial x_k} \quad \text{and} \quad \Gamma(x)^{-1}_{ik} = \frac{\partial x_i}{\partial y_k}. \quad (5.17)$$

(Here, as elsewhere in ChaosBook, a repeated index implies summation.) If x is an equilibrium point, $x = f(x)$, Γ is the matrix inverse of Γ^{-1} , and (5.16) is a *similarity* transformation and thus preserves eigenvalues. It is easy to verify that in the case of period n_p cycle $J'_p(y)$ and $J_p(x)$ are again related by a similarity transformation. (Note, though, that this is not true for $J'(x)$ with $r \neq n_p$). As stability of a flow can always be reduced to stability of a Poincaré return map, a Floquet multiplier of any cycle, for a flow or a map in arbitrary dimension, is a metric invariant of the dynamical system.

section A2.3.1

The i th Floquet (multiplier, eigenvector) pair $(\Lambda_i, \mathbf{e}^{(i)})$ are computed from J evaluated at a periodic point x , $J(x) \mathbf{e}^{(i)}(x) = \Lambda_i \mathbf{e}^{(i)}(x)$, $x \in \mathcal{M}_p$. Multiplying by $\Gamma(x)$ from the left, and inserting $\mathbf{1} = \Gamma(x)^{-1} \Gamma(x)$, we find that the J evaluated at $y = h(x)$ has the same Floquet multiplier,

$$J'_p(y) \mathbf{e}^{(i)}(y)' = \Lambda_i \mathbf{e}^{(i)}(y)', \quad (5.18)$$

but with the eigenvector $\mathbf{e}^{(i)}(x)$ mapped to $\mathbf{e}^{(i)}(y)' = \Gamma(x) \mathbf{e}^{(i)}(x)$.

5.5 Stability of return map cycles

(R. Paškauskas and P. Cvitanović)



If a continuous flow periodic orbit p pierces the Poincaré section $\mathcal{P}$ once, the section point is a fixed point of the return map P with stability (4.25)

$$\hat{J}_{ij} = \left(\delta_{ik} - \frac{v_i U_k}{(v \cdot U)} \right) J_{kj}, \quad (5.19)$$

with all primes dropped, as the initial and the final points coincide, $x' = f^T(x) = x$. If the periodic orbit p pierces the Poincaré section n times, the same observation applies to the n th iterate of P .

We have already established in (4.26) that the velocity $v(x)$ is a zero eigenvector of the Poincaré section Floquet matrix, $\hat{J}v = 0$. Consider next $(\Lambda_\alpha, \mathbf{e}^{(\alpha)})$, the full state space α th (eigenvalue, eigenvector) pair (5.9), evaluated at a periodic point on a Poincaré section,

$$J(x) \mathbf{e}^{(\alpha)}(x) = \Lambda_\alpha \mathbf{e}^{(\alpha)}(x), \quad x \in \mathcal{P}. \quad (5.20)$$

Multiplying (5.19) by $\mathbf{e}^{(\alpha)}$ and inserting (5.20), we find that the full state space Floquet matrix and the Poincaré section Floquet matrix $\hat{J}$ have the same Floquet multiplier

$$\hat{J}(x) \hat{\mathbf{e}}^{(\alpha)}(x) = \Lambda_\alpha \hat{\mathbf{e}}^{(\alpha)}(x), \quad x \in \mathcal{P}, \quad (5.21)$$

where $\hat{\mathbf{e}}^{(\alpha)}$ is a projection of the full state space eigenvector onto the Poincaré section:

$$(\hat{\mathbf{e}}^{(\alpha)})_i = \left(\delta_{ik} - \frac{v_i U_k}{(v \cdot U)} \right) (\mathbf{e}^{(\alpha)})_k. \quad (5.22)$$

Hence, $\hat{J}_p$ evaluated on any Poincaré section point along the cycle p has the same set of Floquet multipliers $\{\Lambda_1, \Lambda_2, \dots, \Lambda_d\}$ as the full state space Floquet matrix J_p , except for the marginal unit Floquet multiplier (5.14).

As established in (4.26), due to the continuous symmetry (time invariance), $\hat{J}_p$ is a rank $d-1$ matrix. We shall refer to the rank $[(d-1-N) \times (d-1-N)]$ submatrix with $N-1$ continuous symmetries quotiented out as the *monodromy matrix* M_p (from Greek *mono-* = alone, single, and *dromo* = run, racecourse, meaning a single run around the stadium). Quotienting continuous symmetries is discussed in chapter 12.

5.6 Won't you be my neighbor?



In what follows, our task will be to determine the size of a *neighborhood* of $x(t)$, and that is why we care about the Floquet multipliers, and especially the unstable (expanding) ones.

Nearby points aligned along the stable (contracting) directions remain in the neighborhood of the trajectory $x(t) = f^t(x_0)$; the ones to keep an eye on are the points which leave the neighborhood along the unstable directions. All chaos arises from flights along these directions. The sub-volume $|\mathcal{M}_{x_0}| = \prod_i^e \Delta x_i$ of the set of points which get no further away from $f^t(x_0)$ than L , the typical size of the system, is fixed by the condition that $\Delta x_i \Lambda_i = O(L)$ in each expanding direction i . Hence the neighborhood size scales as $|\mathcal{M}_{x_0}| \propto O(L^{d_e})/|\Lambda_p| \propto 1/|\Lambda_p|$, where Λ_p is the product of expanding Floquet multipliers (5.7) only; contracting ones play a secondary role. Discussion of sect. 1.5.1, figure 1.8, and figure 5.1 illustrate intersection of initial volume with its return, and chapters 15 and 21 illustrate the key role that unstable directions play in systematically partitioning the state space of a given dynamical system. The contracting directions are so secondary that even infinitely many of them (for example, the infinity of contracting eigen-directions of the spatiotemporal dynamics of Chapter 30) will not matter.

So the dynamically important information is carried by the expanding sub-volume, not the total volume computed so easily in (4.29). That is also the reason why dissipative and Hamiltonian chaotic flows are more alike than one would naively expect for ‘compressible’ vs. ‘incompressible’ flows. In hyperbolic systems, what matters are the expanding directions. Whether the contracting eigenvalues are inverses of the expanding ones or not is of secondary importance. As long as the number of unstable directions is finite, the same theory applies both to the finite-dimensional ODEs and infinite-dimensional PDEs.

Résumé

Periodic orbits play a central role in any invariant characterization of the dynamics, because (a) their existence and inter-relations are a *topological*, coordinate-independent property of the dynamics, and (b) their Floquet multipliers are *metric invariants*, i.e. Floquet multipliers of a periodic orbit remain invariant under any smooth nonlinear change of coordinates $f \rightarrow h \circ f \circ h^{-1}$. Let us summarize the linearized flow notation used throughout ChaosBook.

Stability of invariant solutions: The linear stability of an equilibrium $v(x_q) = 0$ is described by the eigenvalues and eigenvectors $\{\lambda^{(j)}, \mathbf{e}^{(j)}\}$ of the stability matrix A evaluated at the equilibrium point, and the linear stability of a periodic orbit $f^T(x) = x, x \in \mathcal{M}_p$,

$$J_p(x) \mathbf{e}^{(j)}(x) = \Lambda_j \mathbf{e}^{(j)}(x), \quad \Lambda_j = \sigma^{(j)} e^{\lambda^{(j)} T},$$

by its Floquet multipliers, vectors and exponents $\{\Lambda_j, \mathbf{e}^{(j)}\}$, where $\lambda^{(j)} = \mu^{(j)} \pm i\omega^{(j)}$. For every continuous symmetry there is a marginal eigen-direction, with $\Lambda_j = 1$, $\lambda^{(j)} = 0$. With all $1 + N$ continuous symmetries quotiented out (Poincaré sections for time, slices for continuous symmetries of dynamics, see chapter 13), linear stability of a periodic orbit (and, more generally, of a partially hyperbolic torus) is described by the $[(d-1-N) \times (d-1-N)]$ monodromy matrix, all of whose Floquet

multipliers $|\Lambda_j| \neq 1$ are generically strictly hyperbolic,

$$M_p(x) \mathbf{e}^{(j)}(x) = \Lambda_j \mathbf{e}^{(j)}(x), \quad x \in M_p/G.$$

We shall show in chapter 14 that extending the linearized stability hyperbolic eigen-directions into stable and unstable manifolds yields important global information about the topological organization of state space. What matters most are the expanding directions. The physically important information is carried by the unstable manifold, and the expanding sub-volume characterized by the product of expanding Floquet multipliers of J_p . As long as the number of unstable directions is finite, the theory developed here can be applied to flows of arbitrarily high dimension.



in depth:
appendix A4, p. 922



fast track:
chapter 11, p. 192

Commentary

Remark 5.1. Periodic orbits vs. ‘cycles’. Throughout this text, the terms ‘periodic orbit’ and ‘cycle’ (which has many other uses in mathematics) are used interchangeably; while ‘periodic orbit’ is more precise, ‘pseudo-cycle’ is easier on the ear than ‘pseudo-periodic-orbit’. In Soviet times obscure abbreviations were a rage, but here we shy away from acronyms such as UPOs (Unstable Periodic Orbits). How Kadanoff and Tang [7] felt about the matter they let on by referring to these as ‘repulsive cycles’. We refer to unstable periodic orbits simply as ‘periodic orbits’, and the stable ones as ‘limit cycles’. Strogatz [11] refers to periodic orbits as ‘closed orbits’ if they are isolated, in order to distinguish limit cycles and unstable orbits from the continuous family of the harmonic oscillator periodic orbits. Lost in the mists of time is the excitement experienced by the first physicist to discover that there are periodic orbits other than the limit cycles reached by mindless computation forward in time; but once one understands that there are at most several stable limit cycles (SPOs?) as opposed to the Smale horseshoe infinities of unstable cycles (UPOs?), what is gained by prefix ‘U’? A bit like calling all bicycles ‘unstable bicycles’.

Remark 5.2. Periodic orbits and Floquet theory. Study of time-dependent and T -periodic vector fields is a classical subject in the theory of differential equations [6]. The fundamental G. Floquet theorem [2] is from 1883, but stability of periodic orbits was already well understood by G. W. Hill [5] in 1877. G. W. Hill’s work on lunar motions is discussed by M. C. Gutzwiller [3], whose night job for many years were precise calculations of lunar dynamics. In physics literature, Floquet exponents often assume different names according to the context where the theory is applied. They are called Bloch phases in the discussion of Schrödinger equation with a periodic potential [1], or quasi-momenta in the quantum theory of time-periodic Hamiltonians. For clear discussions of stabilities of periodic orbits, see Hale [4] and Robinson [10]. Here a discussion of Floquet theory is given in appendix A4.2.1. For further reading on periodic orbits, consult Moehlis and K. Josić [9] [Scholarpedia.org](https://scholarpedia.org) article.

Question 5.1. Alice

Q In my 61,506-dimensional computation of a Navier-Stokes equilibrium I generated about 30 eigenvectors before I wanted to move on. How many of these eigenvectors are worth generating for a particular solution and why?

A A rule of the thumb is that you need all equilibrium eigenvalues / periodic orbit Floquet exponents with positive real parts, and at least those negative exponents whose magnitude is less or comparable to the largest expanding eigenvalue. More precisely, keep adding the next least contracting eigenvalue to the sum of the preceding ones as long as the sum is positive (Kaplan-Yorke criterion) [8]. Then, just to be conservative, double the number of eigenvalues you keep. You do not need to worry about the remaining (let's say, 100 thousand, or a few million!) eigen-directions for which the negative eigenvalues are of larger magnitude, as they always contract. Nonlinear terms cannot mix them up in such a way that expansion in some directions overwhelms the strongly contracting ones.

chapter 30

Question 5.2. Alice

Q Assume you have a periodic orbit of period T . The Floquet eigenvectors of a Floquet matrix evolved along a periodic orbit are covariant vectors (A4.29). Assume the Jacobian matrix evolved along the periodic orbit has all contracting multipliers, except one multiplier that becomes expanding at $t = T/2$ but then turns contracting at $t = T$. Can we say that the periodic orbit is stable? According to Floquet theory, the periodic orbit is unstable if the multipliers of the Floquet matrix (the Jacobian matrix at time $t = T$) are unstable. When a multiplier of the Jacobian matrix is expanding in the middle of the periodic orbit, at $t = T/2$, what is happening there?

appendix A4.2.1

A The only thing that is meaningful are Floquet multipliers, the eigenvalues of a Floquet matrix (the Jacobian matrix evaluated at the full period T , or its multiple). Floquet eigenvectors are 'covariant' but their orientation depends on the choice of coordinates and where you evaluate them along the orbit. What happens along the orbit is largely irrelevant – intermediate multipliers for times other than the exact period could be contracting or expanding, but that has no invariant meaning. The multipliers depend on coordinate choice and can be changed by time and state space coordinates reparametrizations. For example, consider the Brusselator limit cycle in “Noise is your friend” slides. The cloud of points around the limit cycle shrinks or expands according to the instantaneous stability, but its Floquet multipliers are an invariant property of the limit cycle.

Re “largely irrelevant” above: In “Dimension of turbulence” we argue that it is the angles between covariant vectors approaching arbitrarily close to zero that leads to mixing together the physical dimensions of an attractor, but we have not really established that.

References

- [1] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Holt, Rinehart and Winston, 1976).
- [2] G. Floquet, “Sur les équations différentielles linéaires à coefficients périodiques”, *Ann. Sci. Ec. Norm. Sup.* **12**, 47–88 (1883).
- [3] M. C. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer, New York, 1990).
- [4] J. K. Hale, *Ordinary Differential Equations* (Dover, New York, 1969).

- [5] G. W. Hill, “On the part of the motion of the lunar perigee which is a function of the mean motions of the sun and moon”, *Acta Math.* **8**, 1–36 (1886).
- [6] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
- [7] L. Kadanoff and C. Tang, “Escape rate from strange repellers”, *Proc. Natl. Acad. Sci. USA* **81**, 1276–1279 (1984).
- [8] J. L. Kaplan and J. A. Yorke, “Chaotic behavior of multidimensional difference equations”, in *Functional Differential Equations and Approximation of Fixed Points*, Vol. 730, edited by H.-O. Peitgen and H.-O. Walther, *Lect. Notes Math.* (Springer, Berlin, 1979), pp. 204–227.
- [9] J. Moehlis, K. Josić, and E. T. Shea-Brown, “Periodic orbit”, *Scholarpedia* **1**, 1358 (2006).
- [10] C. Robinson, *Dynamical Systems: Stability, Symbolic Dynamics, and Chaos* (CRC Press, Boca Raton FL, 1995).
- [11] S. H. Strogatz, *Nonlinear Dynamics and Chaos* (Westview Press, Boulder, CO, 2014).

5.7 Examples

The reader is urged to study the examples collected here. To return back to the main text, click on [click to return] pointer on the margin.

Example 5.1. Stability of cycles of 1-dimensional maps. The stability of a prime cycle p of a 1-dimensional map follows from the chain rule (4.43) for stability of the n_p th iterate of the map

$$\Lambda_p = \frac{d}{dx_0} f^{n_p}(x_0) = \prod_{m=0}^{n_p-1} f'(x_m), \quad x_m = f^m(x_0). \quad (5.23)$$

Λ_p is a property of the cycle, not the initial periodic point, as taking any periodic point in the p cycle as the initial one yields the same Λ_p .

A *critical point* x_c is a value of x for which the mapping $f(x)$ has a vanishing derivative, $f'(x_c) = 0$. A periodic orbit of a 1-dimensional map is *stable* if

$$|\Lambda_p| = |f'(x_{n_p})f'(x_{n_p-1}) \cdots f'(x_2)f'(x_1)| < 1,$$

and *superstable* if the orbit includes a critical point, so that the above product vanishes. For a stable periodic orbit of period n , the slope Λ_p of the n th iterate $f^n(x)$ evaluated on a periodic point x (fixed point of the n th iterate) lies between -1 and 1 . If $|\Lambda_p| > 1$, the p -cycle is *unstable*.

[return to p. 113](#)

Example 5.2. Stability of cycles for maps. No matter what method one uses to determine unstable cycles, the theory developed here requires that their Floquet multipliers be evaluated as well. For maps a Floquet matrix is easily evaluated by picking any periodic point as a starting point, running once around a prime cycle, and multiplying the individual periodic point Jacobian matrices according to (4.22). For example, the Floquet matrix M_p for a prime cycle p of period n_p of the Hénon map (3.18) is given by (4.44),

$$M_p(x_0) = \prod_{k=n_p}^1 \begin{bmatrix} -2ax_k & b \\ 1 & 0 \end{bmatrix}, \quad x_k \in \mathcal{M}_p,$$

and the Floquet matrix M_p for a 2-dimensional billiard prime cycle p of period n_p is

$$M_p = (-1)^{n_p} \prod_{k=n_p}^1 \begin{bmatrix} 1 & \tau_k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ r_k & 1 \end{bmatrix}$$

and follows from (9.10). The decreasing order of indices for the products in the above formulas is a reminder that the successive time steps correspond to multiplication from the left, $M_p(x_1) = M(x_{n_p}) \cdots M(x_1)$. We shall compute Floquet multipliers of Hénon map cycles once we learn how to find their periodic orbits, see exercise 16.11.

[return to p. 113](#)

5.8 Exercises

- 5.1. **A limit cycle with analytic Floquet exponent.** There are only two examples of nonlinear flows for which the general solution $(q(t), p(t))$ and the Floquet multipliers can be evaluated analytically. Both are cheats. One example is the 2-dimensional flow

$$\begin{aligned}\dot{q} &= p + q(1 - q^2 - p^2) \\ \dot{p} &= -q + p(1 - q^2 - p^2).\end{aligned}$$

Determine all periodic solutions of this flow, and determine analytically their Floquet exponents. Hint: go to polar coordinates $(q, p) = (r \cos \theta, r \sin \theta)$.

G. Bard Ermentrout

- 5.2. **The other example of a limit cycle with analytic Floquet exponent.** What is the other example of a nonlinear flow for which the Floquet multipliers can be evaluated analytically? Hint: email G.B. Ermentrout.
- 5.3. **Yet another example of a limit cycle with analytic Floquet exponent.** Prove G.B. Ermentrout wrong by solving a third example (or more) of a nonlinear flow for which the Floquet multipliers can be evaluated analytically.

Chapter 6

Lyapunov exponents

[...] people should be taught linear algebra a lot earlier than they are now, because it short-circuits a lot of really stupid and painful and idiotic material.

— Stephen Boyd

LET US APPLY OUR newly acquired tools to the fundamental diagnostics in dynamics: Is a given system ‘chaotic’? And if so, how chaotic? If all points in a neighborhood of a trajectory converge toward the same orbit, the attractor is a fixed point or a limit cycle. However, if the attractor is strange, any two trajectories $x(t) = f^t(x_0)$ and $x(t) + \delta x(t) = f^t(x_0 + \delta x_0)$ that start out very close to each other separate exponentially with time, and in a finite time their separation attains the size of the accessible state space.

example 2.3

section 1.3.1

remark 6.1

This *sensitivity to initial conditions* can be quantified as

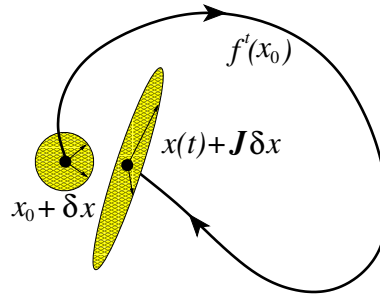
$$\|\delta x(t)\| \approx e^{\lambda t} \|\delta x_0\| \quad (6.1)$$

where λ , the mean rate of separation of trajectories of the system, is called the leading *Lyapunov exponent*. In the limit of infinite time, the Lyapunov exponent is a global measure of the rate at which nearby trajectories diverge, averaged over the strange attractor. As it so often goes with easy ideas, it turns out that Lyapunov exponents are not natural for studying dynamics, and we would have passed them over in silence, were it not for so much literature that talks about them. So in a textbook we are duty bound to explain what all the excitement is about. But then we round the chapter off with a scholarly remark almost as long as the chapter itself: we do not recommend that you evaluate Lyapunov exponents and Lyapunov singular vectors. Instead, compute the stability exponents / covariant vectors.



example 6.1
p. 133

Figure 6.1: The linearized flow maps a swarm of initial points in an infinitesimal spherical neighborhood of squared radius δx^2 at x_0 into an ellipsoid $\delta x^\top (J^\top J) \delta x$ at $x(t)$ a finite time t later, rotated and stretched/compressed along the principal axes by stretches $\{\sigma_j\}$.



6.1 Stretch, strain and twirl

Diagonalizing the matrix: that's the key to the whole thing.

— Governor Arnold Schwarzenegger

In general the Jacobian matrix J is neither diagonal, nor diagonalizable, nor constant along the trajectory. What is the geometrical meaning of mapping a neighborhood by J ? Here insights into continuum mechanics are helpful, in particular polar decomposition visualizes flow linearization as a mapping of the initial ball into an ellipsoid (figure 6.1).



First, a few definitions: A symmetric $[d \times d]$ matrix Q is *positive definite*, $Q > 0$, if $x^\top Q x > 0$ for any nonzero vector $x \in \mathbb{R}^d$. Q is *negative definite*, $Q < 0$, if $x^\top Q x < 0$ for any nonzero vector x . Alternatively, Q is a positive (negative) definite matrix if all its eigenvalues are positive (negative). A matrix R is orthogonal if $R^\top R = \mathbf{1}$, and proper orthogonal if $\det R = +1$. Here the superscript $\top$ denotes the transpose. For example, $(x_1, \dots, x_d)$ is a row vector, $(x_1, \dots, x_d)^\top$ is a column vector. The singular values of matrix J are the square roots of the eigenvalues $\{\sigma_j^2\}$ of $JJ^\dagger$, where † denotes Hermitian transpose.

By the polar decomposition theorem, a deformation J can be factored into a rotation R and a right / left stretch tensor U / V ,

remark 6.2

$$J = RU = VR, \quad (6.2)$$

where R is a proper-orthogonal matrix and U, V are symmetric positive definite matrices with strictly positive real eigenvalues $\{\sigma_1, \sigma_2, \dots, \sigma_d\}$ called *principal stretches* (singular values, Hankel singular values), and with orthonormal eigenvector bases,

$$\begin{aligned} U u^{(i)} &= \sigma_i u^{(i)}, & \{u^{(1)}, u^{(2)}, \dots, u^{(d)}\} \\ V v^{(i)} &= \sigma_i v^{(i)}, & \{v^{(1)}, v^{(2)}, \dots, v^{(d)}\}. \end{aligned} \quad (6.3)$$

If $\sigma_i > 1$, σ_i represents stretching, and if $0 < \sigma_i < 1$, σ_i represents compression along the direction $u^{(i)}$ or $v^{(i)}$. The eigenvectors $\{u^{(j)}\}$ represent the *principal axes of strain* at the initial point x_0 ; whereas, the eigenvectors $\{v^{(j)}\}$ represent the principal axes of strain at the present placement x . From a geometric point of view, J maps the unit sphere into an ellipsoid, figure 6.1; the principal stretches are then

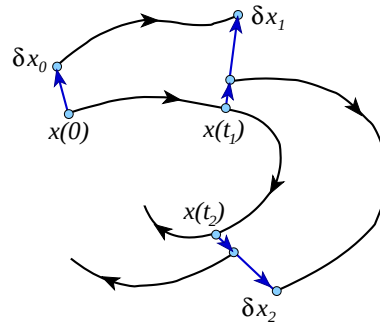


Figure 6.2: A long-time numerical calculation of the leading Lyapunov exponent requires rescaling the distance in order to keep the nearby trajectory separation within the linearized flow range.

the lengths of the semiaxes of this ellipsoid. The rotation matrix R carries the initial axes of strain into the present ones, $V = RUR^T$. The eigenvalues of the

remark 6.1

$$\begin{aligned} \text{right Cauchy-Green strain tensor:} & \quad J^T J = U^2 \\ \text{left Cauchy-Green strain tensor:} & \quad J J^T = V^2 \end{aligned} \quad (6.4)$$

are $\{\sigma_j^2\}$, the squares of principal stretches.



example 6.2
p. 134

6.2 Lyapunov exponents

(J. Mathiesen and P. Cvitanović)

The mean growth rate of the distance $\|\delta x(t)\| / \|\delta x_0\|$ between neighboring trajectories (6.1) is given by the leading *Lyapunov exponent*, which can be estimated for long (but not too long) time t as

$$\lambda \simeq \frac{1}{t} \ln \frac{\|\delta x(t)\|}{\|\delta x(0)\|}. \quad (6.5)$$

For notational brevity, we shall often suppress the dependence of quantities such as $\lambda = \lambda(x_0, t)$, $\delta x(t) = \delta x(x_0, t)$ on the initial point x_0 . One can use (6.5) as is, take a small initial separation δx_0 , track the distance between two nearby trajectories until $\|\delta x(t_1)\|$ gets significantly big, then record $t_1 \lambda_1 = \ln(\|\delta x(t_1)\| / \|\delta x_0\|)$, rescale $\delta x(t_1)$ by factor $\delta x_0 / \delta x(t_1)$, and continue add infinitum, as in figure 6.2, with the leading Lyapunov exponent given by

$$\lambda = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_i t_i \lambda_i, \quad t = \sum_i t_i. \quad (6.6)$$

Deciding what is a safe 'linear range', the distance beyond which the separation vector $\delta x(t)$ should be rescaled, is a dark art.

We can start out with a small δx and try to estimate the leading Lyapunov exponent λ from (6.6), but now that we have quantified the notion of linear stability

in chapter 4, we can do better. The problem with measuring the growth rate of the distance between two points is that as the points separate, the measurement is less and less a local measurement. In the study of experimental time series this might be the only option, but if we have equations of motion, a better way is to measure the growth rate of vectors transverse to a given orbit.

Given the equations of motion, for infinitesimal δx we know the $\delta x_i(t)/\delta x_j(0)$ ratio exactly, and by definition this is the Jacobian matrix

$$\lim_{\delta x(0) \rightarrow 0} \frac{\delta x_i(t)}{\delta x_j(0)} = \frac{\partial x_i(t)}{\partial x_j(0)} = J_{ij}^t(x_0),$$

so the leading Lyapunov exponent can be computed from linearization (4.16)

$$\lambda(x_0) = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \frac{\|J^t(x_0) \delta x_0\|}{\|\delta x_0\|} = \lim_{t \rightarrow \infty} \frac{1}{2t} \ln (\hat{n}^\top J^{t\top} J^t \hat{n}). \quad (6.7)$$

In this formula the scale of the initial separation drops out, only its orientation given by the initial orientation unit vector $\hat{n} = \delta x_0 / \|\delta x_0\|$ matters. If one does not care about the orientation of the separation vector between a trajectory and its perturbation, but only its magnitude, one can interpret $\|J^t \delta x_0\|^2 = \delta x_0^\top (J^{t\top} J^t) \delta x_0$, as the *error correlation matrix*. In the continuum mechanics language, the right Cauchy-Green strain tensor $J^\top J$ (6.4) is the natural object to describe how linearized neighborhoods deform. In the theory of dynamical systems, the *stretches* of continuum mechanics are called the *finite-time Lyapunov* or *characteristic* exponents,

$$\lambda(x_0, \hat{n}; t) = \frac{1}{t} \ln \|J^t \hat{n}\| = \frac{1}{2t} \ln (\hat{n}^\top J^{t\top} J^t \hat{n}). \quad (6.8)$$

They depend on the initial point x_0 and on the direction of the unit vector $\hat{n}$, $\|\hat{n}\| = 1$ at the initial time. If this vector is aligned along the i th principal stretch, $\hat{n} = u^{(i)}$, then the corresponding finite-time Lyapunov exponent (rate of stretching) is given by

$$\lambda_j(x_0; t) = \lambda(x_0, u^{(j)}; t) = \frac{1}{t} \ln \sigma_j(x_0; t). \quad (6.9)$$

We do not need to compute the strain tensor eigenbasis to determine the leading *Lyapunov exponent*,

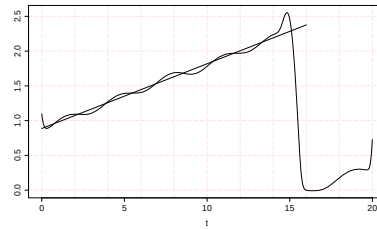
$$\lambda(x_0, \hat{n}) = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \|J^t \hat{n}\| = \lim_{t \rightarrow \infty} \frac{1}{2t} \ln (\hat{n}^\top J^{t\top} J^t \hat{n}). \quad (6.10)$$

By expanding the initial orientation in the strain tensor eigenbasis (6.3), $\hat{n} = \sum (\hat{n} \cdot u^{(i)}) u^{(i)}$, we have

$$\hat{n}^\top J^{t\top} J^t \hat{n} = \sum_{i=1}^d (\hat{n} \cdot u^{(i)})^2 \sigma_i^2 = (\hat{n} \cdot u^{(1)})^2 \sigma_1^2 (1 + O(\sigma_2^2/\sigma_1^2)), \quad (6.11)$$

with stretches ordered by decreasing magnitude, $\sigma_1 > \sigma_2 \geq \sigma_3 \dots$. For long times the largest stretch dominates exponentially in (6.10), provided the orientation $\hat{n}$ of the initial separation was not chosen perpendicular to the dominant

Figure 6.3: A numerical computation of the logarithm of the stretch $\hat{n}^\top (J'^\top J') \hat{n}$ in formula (6.10) for the Rössler flow (2.30), plotted as a function of the Rössler time units. The slope is the leading Lyapunov exponent $\lambda \approx 0.09$. The exponent is positive, so numerics lends credence to the hypothesis that the Rössler attractor is chaotic. The big unexplained jump illustrates perils of Lyapunov exponents numerics. (J. Mathiesen)



expanding eigen-direction $u^{(1)}$. Furthermore, for long times, $J^t \hat{n}$ is dominated by the largest stability multiplier Λ_1 , so the leading Lyapunov exponent is

$$\begin{aligned} \lambda(x_0) &= \lim_{t \rightarrow \infty} \frac{1}{t} \left\{ \ln \left\| \hat{n} \cdot e^{(1)} \right\| + \ln |\Lambda_1(x_0, t)| + O(e^{-2(\lambda_1 - \lambda_2)t}) \right\} \\ &= \lim_{t \rightarrow \infty} \frac{1}{t} \ln |\Lambda_1(x_0, t)|, \end{aligned} \quad (6.12)$$

where $\Lambda_1(x_0, t)$ is the leading eigenvalue of $J^t(x_0)$. The leading Lyapunov exponent now follows from the Jacobian matrix by numerical integration of (4.10). The equations can be integrated accurately for a finite time. Hence, the infinite time limit of (6.7) can be only estimated from a finite set of evaluations of $\frac{1}{2} \ln(\hat{n}^\top J'^\top J' \hat{n})$ as function of time, such as figure 6.3 for the Rössler flow (2.30).

As the local expansion and contraction rates vary along the flow, the temporal dependence exhibits small and large humps. The sudden fall to a low value in figure 6.3 is caused by a close passage to a folding point of the attractor. This evaluation of the Lyapunov exponent proves that the very existence of a strange attractor is a difficult problem. You can use (at your own peril) the approximately monotone part of the curve to estimate the leading Lyapunov exponent by a straight line fit.

As we can already see, we are courting difficulties if we try to calculate the Lyapunov exponent by using the definition (6.12) directly. First of all, the state space is dense with atypical trajectories; for example, if x_0 happens to lie on a periodic orbit p , λ would be simply $\ln |\sigma_{p,1}|/T_p$, a local property of cycle p , not a global property of the dynamical system. Furthermore, even if x_0 happens to be a ‘generic’ state space point, it is still not obvious that $\ln |\sigma_{p,1}(x_0, t)|/t$ should be converging to anything in particular. In a Hamiltonian system with coexisting elliptic islands and chaotic regions, a chaotic trajectory gets captured in the neighborhood of an elliptic island every so often and can stay there for an arbitrarily long time. During such an episode, the orbit is nearly stable and $\ln |\sigma_{p,1}(x_0, t)|/t$ can dip arbitrarily close to 0^+ . For state space volume non-preserving flows, the trajectory can traverse locally contracting regions, and $\ln |\sigma_{p,1}(x_0, t)|/t$ can occasionally go negative; even worse, one never knows whether the asymptotic attractor is periodic or ‘chaotic’, so any finite time estimate of λ might be dead wrong.

exercise 6.3

Résumé

Let us summarize the ‘stability’ chapters 4 to 6. A neighborhood of a trajectory deforms as it is transported by a flow. In the linear approximation, the stability matrix A describes the shearing / compression / expansion of an infinitesimal neighborhood in an infinitesimal time step. The deformation after a finite time t is described by the Jacobian matrix J^t , whose eigenvalues (stability multipliers) depend on the choice of coordinates.

Floquet multipliers and *eigen-vectors* are intrinsic, invariant properties of finite-time, compact invariant solutions, such as periodic orbits and relative periodic orbits; they are explained in chapter 5. *Stability exponents* [7] are the corresponding long-time limits estimated from typical ergodic trajectories.

Finite-time Lyapunov exponents and the associated *principal axes* are defined in (6.8). Oseledec *Lyapunov exponents* are the $t \rightarrow \infty$ limit of these.

Commentary

Remark 6.1. Lyapunov exponents are uncool, and ChaosBook resolutely does not use them at all, for reasons to be explained below.

Eigenvectors / eigenvalues are suitable for studying iterated forms of a matrix, such as the Jacobian matrix J^t or exponential $\exp(tA)$, and are thus a natural tool for studying dynamics. However, principal vectors are not useful for studying dynamics, but they are well-suited for studying the matrix J^t itself. The polar (singular value) decomposition is convenient for numerical work (any matrix, square or rectangular, can be brought to such form), as a way of estimating the effective rank of matrix J by separating the large, significant singular values from the small, negligible singular values.

Lorenz [13, 14, 30] pioneered the use of singular vectors in chaotic dynamics. We found the Goldhirsch, Sulem and Orszag [7] exposition very clear, and we also enjoyed Hoover and Hoover’s [10] pedagogical introduction to computation of Lyapunov spectra by the method of Lagrange multipliers. Greene and Kim [8] discuss singular values vs. Jacobian matrix eigenvalues. While they conclude that “singular values, rather than eigenvalues, are the appropriate quantities to consider when studying chaotic systems,” we beg to differ. Their Fig. 3, which illustrates various semiaxes of the ellipsoid in the case of the Lorenz attractor, as well as the figures in ref. [26], are a persuasive argument for *not* using singular values. The covariant vectors are tangent to the attractor, while the principal axes of strain point away from it. It is the perturbations within the attractor that describe the long-time dynamics; these perturbations lie within the subspace spanned by the leading covariant vectors.

That is the first problem with Lyapunov exponents: stretches $\{\sigma_j\}$ are *not related* to the Jacobian matrix J^t eigenvalues $\{\Lambda_j\}$ in any simple way. The eigenvectors $\{u^{(j)}\}$ of strain tensor $J^T J$, which determine the orientation of the principal axes, are distinct from the Jacobian matrix eigenvectors $\{e^{(j)}\}$. The strain tensor $J^T J$ satisfies no multiplicative semigroup property such as (4.20). Unlike the Jacobian matrix (5.3), the strain tensor $J^{Tr} J^r$ for the r th repeat of a prime cycle p is not given by a power of $J^T J$ for the single traversal of the prime cycle p . Under time evolution the covariant vectors map forward

as $\mathbf{e}^{(j)} \rightarrow J \mathbf{e}^{(j)}$ (transport of the velocity vector (4.9) is an example). In contrast, the principal axes have to be recomputed from scratch for each time t .

If Lyapunov exponents are not dynamical, why are they invoked so frequently? One reason is fear of mathematics. The monumental and therefore rarely read Oseledec [18, 21] Multiplicative Ergodic Theorem states that the limits (6.7–6.12) exist for almost all points x_0 and vectors $\hat{n}$, and that there are at most d distinct Lyapunov exponents $\lambda_i(x_0)$ as $\hat{n}$ ranges over the tangent space. To intimidate the reader further we note in passing that “moreover there is a fibration of the tangent space $\mathbf{T}_x \mathcal{M}$, $L^1(x) \subset L^2(x) \subset \dots \subset L^r(x) = \mathbf{T}_x \mathcal{M}$, such that if $\hat{n} \in L^i(x) \setminus L^{i-1}(x)$ the limit (6.7) equals $\lambda_i(x)$.” The Oseledec proof is important mathematics, but the method is not helpful in elucidating dynamics.

The other reason to study singular vectors is physical and practical. Lorenz [13, 14, 30] was interested in the propagation of errors, i.e., how does a cloud of initial points $x(0) + \delta x(0)$, distributed as a Gaussian with covariance matrix $Q(0) = \langle \delta x(0) \delta x(0)^\top \rangle$, evolve in time? For linearized flow with initial isotropic distribution $Q(0) = \epsilon \mathbf{1}$, the answer is given by the left Cauchy-Green strain tensor,

$$Q(t) = \langle \delta x(0) J J^\top \delta x(0)^\top \rangle = J Q(0) J^\top = \epsilon J J^\top. \quad (6.13)$$

The deep problem with Lyapunov exponents is that the intuitive definition (6.5) depends on the notion of distance $\|\delta x(t)\|$ between two state space points. The Euclidean (or L^2) distance is natural in the theory of 3D continuous media, but what the norm should be for other state spaces is far from clear, especially in high dimensions and for PDEs. As we have shown in sect. 5.3, Floquet multipliers are invariant under all local smooth non-linear coordinate transformations, and they are intrinsic to the flow. Additionally, Floquet eigenvectors are independent of the definition of the norm [26]. In contrast, the stretches $\{\sigma_j\}$, and the right/left principal axes depend on the choice of the norm. Appending them to dynamics destroys its invariance.

In dynamical systems literature, there is probably no name more liberally used and more confusing than that of Lyapunov (AKA Liapunov). Singular values / principal axes of strain tensor $J^\top J$ (objects natural to the theory of deformations) and their long-time limits can indeed be traced back to the thesis of Lyapunov [15, 18] (English translation [16]), and justly deserve sobriquet ‘Lyapunov’. Oseledec [18] refers to them as ‘Liapunov characteristic numbers’, and Eckmann and Ruelle [4] as ‘characteristic exponents’. The natural objects in dynamics are the linearized flow Jacobian matrix J' and its corresponding eigenvalues and eigenvectors (stability or characteristic multipliers and covariant vectors). Why should they also be called ‘Lyapunov’? The Jacobian matrix eigenvectors $\{\mathbf{e}^{(j)}\}$ (the covariant vectors) are often called ‘covariant Lyapunov vectors’, ‘Lyapunov vectors’, or ‘stationary Lyapunov basis’ [5], even though *they are not* the eigenvectors that correspond to the Lyapunov exponents. That’s just confusing, for no good reason. The Lyapunov paper [15] is not about the linear stability Jacobian matrix J , but it is about $J^\top J$ and the associated principal axes. However, Trevisan [26] refers to covariant vectors as ‘Lyapunov vectors’, and Radons [29] calls them ‘Lyapunov modes’, motivated by thinking of these eigenvectors as a generalization of ‘normal modes’ of mechanical systems; whereas, Takeuchi and Chat   [24] define the i th ‘Lyapunov mode’ as $\{\lambda_j, \mathbf{e}^{(j)}\}$, the set of the i th stability exponent and the associated covariant vector. Kunihiro *et al.* [12] call the eigenvalues of stability matrix (4.3), evaluated at a given instant in time, the ‘local Lyapunov exponents’, and they refer to the set of stability exponents (4.8) for a finite time Jacobian matrix as the ‘intermediate Lyapunov exponent’, “averaged” over a finite time period. Then there is the unrelated, but correctly attributed ‘Lyapunov equation’ of control theory, which is the linearization of the ‘Lyapunov function’. Additionally, the term ‘Lyapunov orbit’ is often used in celestial mechanics and is entirely unrelated to any of the objects discussed above.

In short, we do not recommend that you evaluate Lyapunov exponents. Instead, compute stability exponents and the associated covariant vectors. This approach will cost less and offer you more insight. Whatever you call your exponents, please state clearly how they are being computed. While the Lyapunov exponents are a diagnostic tool for chaos, we doubt their ability to predict observables with any physical significance. This is a minority position - in the literature one encounters many provocative speculations, especially in the context of foundations of statistical mechanics ('hydrodynamic' modes) and the existence of a Lyapunov spectrum in the thermodynamic limit of spatiotemporal chaotic systems.

Remark 6.2. Matrix decompositions of the Jacobian matrix. The 'Cartesian decomposition' separates stability matrix A into a symmetric stretching matrix and an antisymmetric spin matrix,

$$D = (A + A^\top)/2, \quad \Omega = (A - A^\top)/2. \quad (6.14)$$

The stretching matrix describes the infinitesimal volume and shape changes, while the spin matrix describes the infinitesimal rigid body rotations. The decomposition into these two non-commuting matrices suggests that one could compute the stretching rates (Lyapunov exponents) of the Jacobian matrix, and the total rotation of the orbit neighborhood by integrating the stability matrix A in the rotating frame (we have not seen this approach implemented in the Lyapunov exponents literature). 'Polar decomposition' of a matrix or linear operator is similar to factoring a complex number into its polar form, $z = r \exp(\phi)$, and is explained in refs. [9, 11, 19, 27]. One can go a step further than polar decomposition (6.2) by diagonalizing the symmetric matrix by a second rotation, which is equivalent to the product of a rotation and a symmetric matrix. Thus, any matrix with real elements can be expressed in singular value decomposition (SVD) form

$$J = R_1 D R_2^\top, \quad (6.15)$$

where D is diagonal and real, and R_1, R_2 are orthogonal matrices, unique up to permutations of rows and columns. The diagonal elements $\{\sigma_1, \sigma_2, \dots, \sigma_d\}$ of D are *singular values* of J .

Though singular value decomposition provides geometrical insights into how tangent dynamics acts, many popular algorithms for asymptotic stability analysis (computing Lyapunov spectrum) employ another standard matrix decomposition, the QR scheme [17], through which a nonsingular matrix J is (uniquely) written as a product of an orthogonal and an upper triangular matrix $J = QR$. This can be thought of as a Gram-Schmidt decomposition of the column vectors of J . The geometric meaning of QR decomposition is that the volume of the d -dimensional parallelepiped spanned by the column vectors of J has a volume coinciding with the product of the diagonal elements of the triangular matrix R , whose role is thus pivotal in algorithms computing Lyapunov spectra [23].

Remark 6.3. Numerical evaluation of Lyapunov exponents. There are volumes of literature on numerical computation of the Lyapunov exponents, see for example refs. [3, 4, 25, 28]. For early numerical methods to compute Lyapunov vectors, see refs. [1, 22]. The drawback of the Gram-Schmidt method is that the vectors so constructed are orthogonal by fiat; whereas, the stable / unstable eigenvectors of the Jacobian matrix are in general not orthogonal. Hence the Gram-Schmidt vectors are not covariant, i.e., the linearized dynamics does not transport them into the eigenvectors of the Jacobian matrix computed further downstream. For computation of covariant vectors, see refs. [6, 20].

References

- [1] G. Benettin, L. Galgani, A. Giorgilli, and J. M. Strelcyn, “Lyapunov characteristic exponents for smooth dynamical systems; a method for computing all of them. Part 1: Theory”, *Meccanica* **15**, 9–20 (1980).
- [2] I. C. Christov, R. M. Lueptow, and J. M. Ottino, “Stretching and folding versus cutting and shuffling: An illustrated perspective on mixing and deformations of continua”, *Amer. J. Phys.* **79**, 359–367 (2011).
- [3] J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, and S. Ciliberto, “Liapunov exponents from time series”, *Phys. Rev. A* **34**, 4971–4979 (1986).
- [4] J.-P. Eckmann and D. Ruelle, “Ergodic theory of chaos and strange attractors”, *Rev. Mod. Phys.* **57**, 617–656 (1985).
- [5] S. V. Ershov and A. B. Potapov, “On the concept of stationary Lyapunov basis”, *Physica D* **118**, 167–198 (1998).
- [6] F. Ginelli, P. Poggi, A. Turchi, H. Chaté, R. Livi, and A. Politi, “Characterizing dynamics with covariant Lyapunov vectors”, *Phys. Rev. Lett.* **99**, 130601 (2007).
- [7] I. Goldhirsch, P. L. Sulem, and S. A. Orszag, “Stability and Lyapunov stability of dynamical systems: A differential approach and a numerical method”, *Physica D* **27**, 311–337 (1987).
- [8] J. M. Greene and J.-S. Kim, “The calculation of Lyapunov spectra”, *Physica D* **24**, 213–225 (1987).
- [9] M. Gurtin, *An Introduction to Continuum Mechanics* (Academic, New York, 1981).
- [10] W. G. Hoover and C. G. Hoover, “Local Gram–Schmidt and covariant Lyapunov vectors and exponents for three harmonic oscillator problems”, *Commun. Nonlinear Sci. Numer. Simul.* **17**, 1043–1054 (2012).
- [11] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge Univ. Press, Cambridge UK, 2012).
- [12] T. Kunihiro, B. Müller, A. Ohnishi, A. Schäfer, T. T. Takahashi, and A. Yamamoto, “Chaotic behavior in classical Yang–Mills dynamics”, *Phys. Rev. D* **82**, 114015 (2010).
- [13] E. N. Lorenz, “A study of the predictability of a 28-variable atmospheric model”, *Tellus* **17**, 321–333 (1965).
- [14] E. N. Lorenz, “Irregularity: a fundamental property of the atmosphere”, *Tellus A* **36**, 98–110 (1984).
- [15] A. Lyapunov, “Problème général de la stabilité du mouvement”, *Ann. Math. Studies* **17**, Russian original Kharkow, 1892, 531–534 (1977).
- [16] A. M. Lyapunov, “The general problem of the stability of motion”, *Int. J. Control* **55**, 531–534 (1992).
- [17] C. Meyer, *Matrix Analysis and Applied Linear Algebra* (SIAM, Philadelphia, 2000).

- [18] V. I. Oseledec, “A multiplicative ergodic theorem. Liapunov characteristic numbers for dynamical systems”, *Trans. Moscow Math. Soc.* **19**, 197–221 (1968).
- [19] J. M. Ottino, *The Kinematics of Mixing: Stretching, Chaos and Transport* (Cambridge Univ. Press, Cambridge UK, 1989).
- [20] A. Politi, A. Torcini, and S. Lepri, “Lyapunov exponents from node-counting arguments”, *J. Phys. IV* **8**, 263–270 (1998).
- [21] M. Pollicott, *Lectures on Ergodic Theory and Pesin Theory on Compact Manifolds* (Cambridge Univ. Press, Cambridge UK, 1993).
- [22] I. Shimada and T. Nagashima, “A numerical approach to ergodic problem of dissipative dynamical systems”, *Progr. Theor. Phys.* **61**, 1605–1616 (1979).
- [23] C. Skokos, “The Lyapunov characteristic exponents and their computation”, in *Dynamics of Small Solar System Bodies and Exoplanets*, edited by J. J. Souchay and R. Dvorak (Springer, New York, 2010), pp. 63–135.
- [24] K. A. Takeuchi and H. Chaté, “Collective Lyapunov modes”, *J. Phys. A* **46**, 254007 (2013).
- [25] J.-L. Thiffeault, “Derivatives and constraints in chaotic flows: asymptotic behaviour and a numerical method”, *Physica D* **172**, 139–161 (2002).
- [26] A. Trevisan and F. Pancotti, “Periodic orbits, Lyapunov vectors, and singular vectors in the Lorenz system”, *J. Atmos. Sci.* **55**, 390 (1998).
- [27] C. A. Truesdell, *A First Course in Rational Continuum Mechanics. General Concepts*, Vol. 1 (Academic, New York, 1977).
- [28] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, “Determining Lyapunov exponents from a time series”, *Physica D* **16**, 285–317 (1985).
- [29] H.-I. Yang and G. Radons, “Comparison between covariant and orthogonal Lyapunov vectors”, *Phys. Rev. E* **82**, 046204 (2010).
- [30] S. Yoden and M. Nomura, “Finite-time Lyapunov stability analysis and its application to atmospheric predictability”, *J. Atmos. Sci.* **50**, 1531–1543 (1993).

6.3 Examples

The reader is urged to study the examples collected here. To return back to the main text, click on [click to return] pointer on the margin.

Example 6.1. Lyapunov exponent. Given a 1-dimensional map, consider observable $\lambda(x) = \ln |f'(x)|$ and Birkhoff sum (for definitions, see sect. 20.1.1)

$$A(x_0, t) = \sum_{k=0}^{n-1} \ln |f'(x_k)| = \ln \left| \prod_{k=0}^{n-1} f'(x_k) \right| = \ln \left| \frac{\partial f^n}{\partial x}(x_0) \right|.$$

The Lyapunov exponent is the average rate of the expansion

$$\lambda(x_0) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \ln |f'(x_k)|.$$

[return to p. 124](#)

Example 6.2. Singular values and geometry of deformations. Suppose we are in three dimensions, and the Jacobian matrix J is not singular (yet another confusing usage of word ‘singular’), so that the diagonal elements of D in (6.15) satisfy $\sigma_1 \geq \sigma_2 \geq \sigma_3 > 0$. Consider how J maps the unit ball $S = \{x \in \mathbb{R}^3 \mid x^2 = 1\}$. V is orthogonal (rotation/reflection), so $V^T S$ is still the unit sphere. Then D maps S onto ellipsoid $\tilde{S} = \{y \in \mathbb{R}^3 \mid y_1^2/\sigma_1^2 + y_2^2/\sigma_2^2 + y_3^2/\sigma_3^2 = 1\}$, whose principal axes directions - y coordinates - are determined by V . Finally the ellipsoid is further rotated by the orthogonal matrix U . The local directions of stretching and their images under J are called the right-hand and left-hand singular vectors for J and are given by the columns in V and U respectively; it is easy to check that $Jv_k = \sigma_k u_k$, if v_k, u_k are the k -th columns of V and U .

[return to p. 126](#)

6.4 Exercises

- 6.1. **Principal stretches.** Consider $dx = f(x_0 + dx_0) - f(x_0)$, and show that $dx = Mdx_0 +$ higher order terms when $\|dx_0\| \ll 1$. (Hint: use Taylor expansion for a vector function.) Here, $\|dx_0\| \equiv \sqrt{dx_0 \cdot dx_0}$ is the norm induced by the usual Euclidean dot (inner) product. Then let $dx_0 = (d\ell)e_i$ and show that $\|dx_0\| = d\ell$ and $\|dx\| = \sigma_i d\ell$. (Christov *et al.* [2])
- 6.2. **Eigenvalues of the Cauchy-Green strain tensor.** Show that $\kappa_i = \sigma_i^2$ using the definition of C , the polar decomposition theorem, and the properties of eigenvalues. (Christov *et al.* [2])
- 6.3. **How unstable is the Hénon attractor?**

- (a) Evaluate numerically the Lyapunov exponent λ by iterating some 100,000 times or so the Hénon map

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 - ax^2 + y \\ bx \end{bmatrix}$$

for $a = 1.4$, $b = 0.3$.

- (b) Would you describe the result as a 'strange attractor'? Why?
- (c) How robust is the Lyapunov exponent for the Hénon attractor? Evaluate numerically the Lyapunov exponent by iterating the Hénon map for $a = 1.39945219$, $b = 0.3$. How much do you now trust your result for part (a) of this exercise?
- (d) Re-examine this computation by plotting the iterates, and erasing the plotted points every 1000 iterates or so. Keep at it until the 'strange' attractor vanishes like the smile of the Cheshire cat. What replaces it? Do a few numerical experiments to estimate the length of a typical transient before the dynamics settles into this long-time attractor.
- (e) Use your Newton search routine to confirm the existence of this attractor. Compute its Lyapunov exponent, compare with your numerical result from above. What is the itinerary of the attractor.
- (f) Would you describe the result as a 'strange attractor'? Do you still have confidence in claims such as the one made for part (b) of this exercise?

6.4. Rössler attractor Lyapunov exponents.

- (a) Evaluate numerically the expanding Lyapunov exponent λ_e of the Rössler attractor (2.30).
- (b) Plot your own version of figure 6.3. Do not worry if it looks different, as long as you understand why your plot looks the way it does. (Remember the nonuniform contraction/expansion of figure 4.3.)
- (c) Give your best estimate of λ_e . The literature gives surprisingly inaccurate estimates - see whether you can do better.
- (d) Estimate the contracting Lyapunov exponent λ_c . Even though it is much smaller than λ_e , a glance at the stability matrix (4.31) suggests that you can probably get it by integrating the infinitesimal volume along a long-time trajectory, as in (4.29).

Chapter 7

Fixed points

SO FAR WE HAVE LEARNED that periodic orbits offer invariant characterization of dynamics in two ways: (a) their existence and inter-relations are a topological, coordinate-independent property of the dynamics, and (b) their Floquet multipliers form an infinite set of *metric invariants*. Typically they are unstable and hard to find. But do we really need them? By chapter 21 you will understand that the answer is a resounding yes.

Sadly, searching for periodic orbits will never become as popular as a week on Côte d’Azur, or publishing yet another log-log plot in *Phys. Rev. Letters*. This chapter is one of four hands-on chapters on extraction of periodic orbits, and can be skipped on first reading - you can return to it whenever the need for finding actual cycles arises.



fast track:

chapter 8, p. 146

A serious cyclist will ask “Where are the cycles? And what if they are long?” and read chapter 16. She will want to also learn about the variational methods which will enable her to find arbitrarily long, arbitrarily unstable cycles, and read chapter 34. So here is the key and unavoidable numerical task we must face up to: find “all(?)” solutions (x, T) , $x \in \mathbb{R}^d$, $T \in \mathbb{R}_+$ satisfying the *periodic orbit condition*

chapter 16

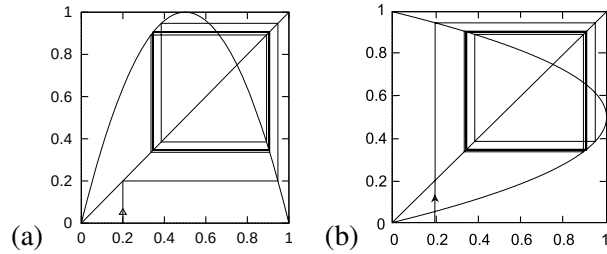
chapter 34

$$\begin{aligned} f^T(x) &= x, & T > 0, & \quad (\text{flow}) \\ f^n(x) &= x, & n \geq 1, & \quad (\text{map}) \end{aligned} \tag{7.1}$$

for a given flow or map.

A *prime cycle* p of period T_p is a single traversal of the periodic orbit, so our task will be to find a periodic point $x \in \mathcal{M}_p$ and the shortest time T_p for which

Figure 7.1: (a) The inverse time transient to the $\overline{01}$ -cycle of the Ulam logistic map $f(x) = 4x(1-x)$ from an initial guess $x = 0.2$. (b) The same dynamics, but now plotted as the forward iteration of the doubly-valued inverse map $f^{-1}(x)$. At each iteration we chose the 0 (respectively 1) branch. For $f(x)$, the $\overline{01}$ -cycle is an unstable cycle; for $f^{-1}(x)$ it is a stable, attracting cycle.



(7.1) has a solution. A periodic point of a flow f^t crossing a Poincaré section n times is a fixed point of P^n , the n th iterate of the return map P (3.1); hence, we shall refer to all cycles as “fixed points” in this chapter. By cyclic invariance, Floquet multipliers and the period of the cycle are independent of the choice of the initial point, so it will suffice to solve (7.1) at a single periodic point. section 5.3

If the cycle is an attracting limit cycle with a sizable basin of attraction, it can be found by integrating the flow for a sufficiently long time. If the cycle is unstable, simple integration forward in time will not reveal it, and the methods to be described here need to be deployed. In essence, any method for finding a cycle is based on devising a new dynamical system which possesses the same cycle, but for which this cycle is attractive. Beyond that, there is great freedom in constructing such systems, and many different methods are used in practice.

7.1 One-dimensional maps

(F. Christiansen)

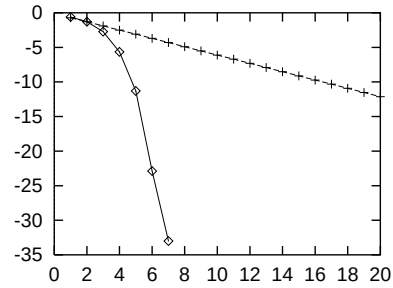
So far we have given some qualitative hints for how to set out on a periodic orbit hunt. In what follows, we teach you how to nail down periodic orbits numerically.

7.1.1 Inverse iteration

Let us first consider a very simple method to find the unstable cycles of a 1-dimensional map such as the logistic map. Unstable cycles of 1-dimensional maps are attracting cycles of the inverse map. The inverse map is not single-valued, so at each backward iteration we have a choice of branch to make. By choosing the branch according to the symbolic dynamics of the cycle we are trying to find, we will automatically converge to the desired cycle. Figure 7.1 shows such a path to the $\overline{01}$ -cycle of the logistic map. The rate of convergence is given by the stability of the cycle, i.e., the convergence is exponentially fast, see figure 7.2.

The method of inverse iteration is fine for finding cycles for 1-d maps and some 2-dimensional systems such as the repeller of exercise 16.11. It is not particularly fast, however, especially if the inverse map is not known analytically. It

Figure 7.2: Convergence of Newton's method ($\diamond$) vs. inverse iteration (+). The error after n iterations searching for the $\overline{01}$ -cycle of the logistic map $f(x) = 4x(1 - x)$ with an initial starting guess of $x_1 = 0.2, x_2 = 0.8$. The y-axis is $\log_{10}$ of the error. The difference between the exponential convergence of the inverse iteration method and the super-exponential convergence of Newton's method is dramatic.



also completely fails for higher dimensional systems when one encounters both stable and unstable directions. Inverse iteration will exchange them, but we will still be left with both stable and unstable directions. The best strategy is to directly attack the problem of finding solutions of the periodic orbit condition $f^T(x) = x$.

7.1.2 Newton method

John Keats has written, "Beauty is truth, truth beauty" He has also written "A thing of beauty is a joy forever." I wish to add, beauty is simple and it is profound. I hope that my few words will convince you that Newton's Method is a concept of great beauty.

—Stephen Smale, [The Concinnitas Project](#)

Newton's method for determining a zero x^* of a function $F(x)$ of one variable is based on a linearization around a starting guess x_0 :

$$F(x) \approx F(x^{(0)}) + F'(x^{(0)})(x - x^{(0)}). \quad (7.2)$$

An approximate solution $x^{(1)}$ of $F(x) = 0$ is

$$x^{(1)} = x^{(0)} - F(x^{(0)})/F'(x^{(0)}). \quad (7.3)$$

The approximate solution can then be used as a new starting guess in an iterative process. A fixed point of a map f is a solution to $F(x) = x - f(x) = 0$. We determine x by iterating

$$\begin{aligned} x^{(m)} &= g(x^{(m-1)}) = x^{(m-1)} - F(x^{(m-1)})/F'(x^{(m-1)}) \\ &= x^{(m-1)} - \frac{1}{1 - f'(x^{(m-1)})}(x^{(m-1)} - f(x^{(m-1)})). \end{aligned} \quad (7.4)$$

Provided that the fixed point is not marginally stable, $f'(x) \neq 1$ at the fixed point x , a fixed point of f is a super-stable fixed point of the Newton-Raphson map g , $g'(x) = 0$. Also, with a sufficiently good initial guess, the Newton-Raphson iteration will converge super-exponentially fast.

To illustrate the efficiency of Newton's method we compare it to the inverse iteration method in figure 7.2. Newton's method wins hands down; the number

Figure 7.3: Newton's method for finding zeros of functions, $f(x_{\text{zero}}) = 0$, an idealized sketch.

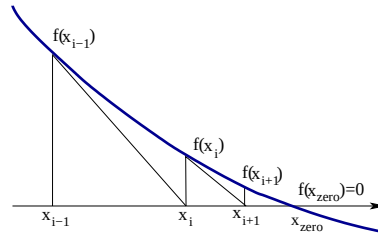
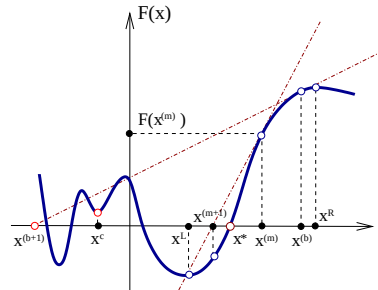


Figure 7.4: Newton's method of figure 7.3 in real life: bad initial guess $x^{(b)}$ leads to the Newton estimate $x^{(b+1)}$ far away from the desired zero of $F(x)$. Sequence $\dots, x^{(m)}, x^{(m+1)}, \dots$, starting with a good guess converges super-exponentially to x^* . The method diverges if it iterates into the basin of attraction of a local minimum x^c .



of significant digits of the accuracy of the x estimate typically doubles with each iteration.

In order to avoid jumping too far from the desired x^* (see figure 7.4), one often initiates the search by the *damped Newton method*,

$$\Delta x^{(m)} = x^{(m+1)} - x^{(m)} = -\frac{F(x^{(m)})}{F'(x^{(m)})} \Delta\tau, \quad 0 < \Delta\tau \leq 1.$$

This method takes small $\Delta\tau$ steps at the beginning, reinstating to the full $\Delta\tau = 1$ jumps only when sufficiently close to the desired x^* .

 example 7.1
p. 142

7.2 Flows

(R. Paškauskas and P. Cvitanović)

For a continuous time flow, the periodic orbit the Floquet multiplier (5.14) along the flow direction always equals unity; the separation of any two points along a cycle remains unchanged after a completion of the cycle. More unit Floquet multipliers arise if the flow satisfies conservation laws, such as the symplectic invariance for Hamiltonian flows, or the dynamics is equivariant under a continuous symmetry transformation.

section 5.3.1

section 12.3

Let us apply Newton's method of (7.3) to search for periodic orbits with unit Floquet multipliers, starting with the case of a *continuous time flow*. Assume that the periodic orbit condition (7.1) holds for $x + \Delta x$ and $T + \Delta t$, with the initial guesses x and T close to the desired solution, i.e., with $|\Delta x|, \Delta t$ small. The Newton



setup (7.3)

$$\begin{aligned} 0 &= x + \Delta x - f^{T+\Delta t}(x + \Delta x) \\ &\approx x - f^T(x) + (1 - J(x)) \cdot \Delta x - v(f^T(x))\Delta t \end{aligned} \quad (7.5)$$

suffers from two shortcomings. First, we now need to solve not only for the periodic point x , but for the period T as well. Second, the marginal, unit Floquet multiplier (5.14) along the flow direction (arising from the time-translation invariance of a periodic orbit) renders the factor $(1 - J)$ in (7.4) non-invertible. If x is close to the solution, $f^T(x) \approx x$, then $J(x) \cdot v(x) = v(f^T(x)) \approx v(x)$. If Δx is parallel to the velocity vector, the derivative term $(1 - J) \cdot \Delta x \approx 0$, and it becomes harder to invert $(1 - J)$ as the iterations approach the solution.

As a periodic orbit p is a 1-dimensional set of points invariant under dynamics, a guess using Newton's method is not improved by picking Δx such that the new point lies on the orbit of the initial one. Hence, we need to constrain the variation Δx to directions transverse to the flow, by requiring, for example, that

$$v(x) \cdot \Delta x = 0. \quad (7.6)$$

Combining this constraint with the variational condition (7.5) we obtain a Newton setup for flows, best displayed in the matrix form:

$$\begin{bmatrix} 1 - J(x) & -v(x) \\ v(x) & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta t \end{pmatrix} = - \begin{pmatrix} x - f(x) \\ 0 \end{pmatrix} \quad (7.7)$$

This illustrates the general strategy for determining periodic orbits in the presence of continuous symmetries. For each symmetry, pick a point on the orbit by imposing a constraint, and compute the value of the corresponding continuous parameter (here the period T) by iterating the enlarged set of Newton's equations. Constraining the variations to transverse ones, thus fixes both of Newton's shortcomings; it breaks the time-translation invariance, and the period T can easily be determined once the fixed point has been found (hence we omit the superscript in f^T for the remainder of this discussion).

More generally, the Poincaré section technique of sect. 3.1 turns the periodic orbit search into a fixed point search on a suitably defined surface of section, with a neighboring point variation Δx with respect to a reference point x constrained to *stay* on the surface manifold (3.2),

$$U(x + \Delta x) = U(x) = 0. \quad (7.8)$$

The price to pay is constraints imposed by the section. In order to *stay* on the surface, arbitrary variation Δx is not allowed.

 example 7.2
p. 143

Résumé

There is no general computational algorithm that is guaranteed to find all solutions (up to a given period T_{max}) to the periodic orbit condition

$$f^{t+T}(x) = f^t(x), \quad T > 0$$

for a general flow or mapping. Due to the exponential divergence of nearby trajectories in chaotic dynamical systems, a direct solution of the periodic orbit condition can be very unstable numerically. With a sufficiently good initial guess for a point x on the cycle, however, the Newton-Raphson formula

$$\begin{pmatrix} 1 - J & -v(x) \\ a & 0 \end{pmatrix} \begin{pmatrix} \delta x \\ \delta T \end{pmatrix} = \begin{pmatrix} f(x) - x \\ 0 \end{pmatrix}$$

yields an improved estimate $x' = x + \delta x, T' = T + \delta T$. Newton-Raphson iteration then yields the period T and the location of a periodic point x_p in the Poincaré section $(x_p - x_0) \cdot a = 0$, where a is a vector normal to the Poincaré section at x_0 .

Commentary

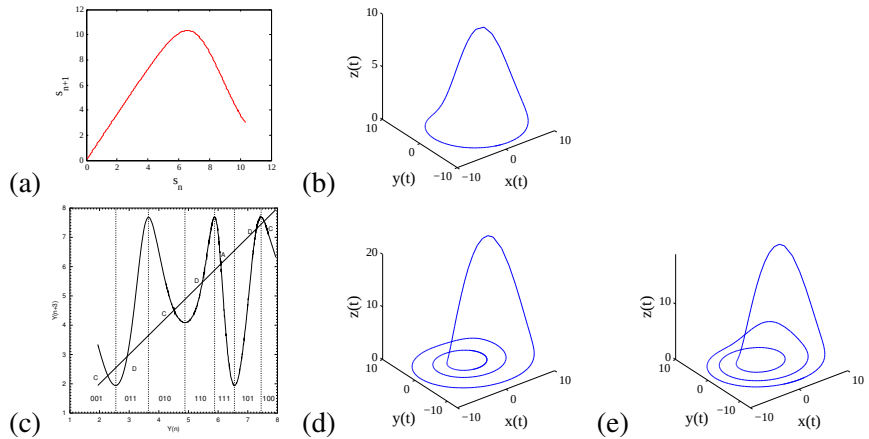
Remark 7.1. Piecewise linear maps. The Lozi map (3.20) is linear, and hundred of thousands of cycles can easily be computed by using $[2 \times 2]$ matrix multiplication and inversion.

Remark 7.2. Newton gone wild. Skowronek and Gora [2] offer an interesting discussion of Newton iterations gone wild while searching for roots of polynomials as simple as $x^2 + 1 = 0$.

References

- [1] C. Dong, “Organization of the periodic orbits in the Rössler flow”, *Int. J. Mod. Phys. B* **22**, 1850227 (2018).
- [2] L. Skowronek and P. F. Gora, “Chaos in Newtonian iterations: Searching for zeros which are not there”, *Acta Phys. Polonica B* **38**, 1909 (2007).

Figure 7.5: (a) $y \rightarrow P_1(y, z)$ return map for $x = 0, y > 0$ Poincaré section of the Rössler flow figure 2.6. (b) The $\bar{1}$ -cycle found by taking the fixed point $y_{k+n} = y_k$ together with the fixed point of the $z \rightarrow z$ return map (not shown) an initial guess $(0, y^{(0)}, z^{(0)})$ for the Newton-Raphson search. (c) $y_{k+3} = P_1^3(y_k, z_k)$, the third iterate of return map (3.1) together with the corresponding plot for $z_{k+3} = P_2^3(y_k, z_k)$, is used to pick starting guesses for the Newton-Raphson searches for the two 3-cycles: (d) the 001 cycle, and (e) the 011 cycle. (A. Basu)



7.3 Examples

Example 7.1. Rössler attractor.

We run a long simulation of the Rössler flow f^t , plot a Poincaré section, as in figure 3.3, and extract the corresponding return map P , as in figure 3.4. Luck is with us, since figure 7.5(a) return map $y \rightarrow P_1(y, z)$ is quite reminiscent of a parabola, we take the unimodal map symbolic dynamics, sect. 14.3, as our guess for the covering dynamics. Strictly speaking, the attractor is “fractal,” but for all practical purposes the return map is 1-dimensional; your printer will need a resolution better than 10^{14} dots per inch to even begin resolving its structure.

Periodic points of a prime cycle p of period n_p for the $x = 0, y > 0$ Poincaré section of the Rössler flow figure 2.6 are fixed points $(y, z) = P^{n_p}(y, z)$ of the n th return map.

By using the fixed point $y_{k+1} = y_k$ in figure 7.5(a) along with the simultaneous fixed point of the $z \rightarrow P_1(y, z)$ return map (not shown), we obtain a starting guess $(0, y^{(0)}, z^{(0)})$ for the Newton-Raphson search for the cycle p with symbolic dynamics label $\bar{1}$. Then we can find the cycle figure 7.5(b) with the Poincaré section point $(0, y_p, z_p)$, period T_p , expanding, marginal, contracting Floquet multipliers $(\Lambda_{p,e}, \Lambda_{p,m}, \Lambda_{p,c})$, and the corresponding Lyapunov exponents $(\lambda_{p,e}, \lambda_{p,m}, \lambda_{p,c})$:

exercise 7.1

$$\begin{aligned}
 \bar{1}\text{-cycle:} \quad (x, y, z) &= (0, 6.09176832, 1.2997319) \\
 T_1 &= 5.88108845586 \\
 (\Lambda_{1,e}, \Lambda_{1,m}, \Lambda_{1,c}) &= (-2.40395353, 1 + 10^{-14}, -1.29 \times 10^{-14}) \\
 (\lambda_{1,e}, \lambda_{1,m}, \lambda_{1,c}) &= (0.149141556, 10^{-14}, -5.44). \quad (7.9)
 \end{aligned}$$

The Newton-Raphson method that we used is described in sect. 7.2.

As an example of a search for longer cycles, we use $y_{k+3} = P_1^3(y_k, z_k)$, the third iterate of the return map (3.1) plotted in figure 7.5(c), together with a corresponding plot for $z_{k+3} = P_2^3(y_k, z_k)$, to pick starting guesses for the Newton-Raphson searches for the two 3-cycles plotted in figure 7.5(d), (e). For a listing of the short cycles of the Rössler flow, consult exercise 7.1.

The numerical evidence suggests (though a proof is lacking) that all cycles that comprise the strange attractor of the Rössler flow are hyperbolic, each with an expanding

eigenvalue $|\Lambda_e| > 1$, a contracting eigenvalue $|\Lambda_c| < 1$, and a marginal eigenvalue $|\Lambda_m| = 1$ corresponding to displacements along the direction of flow.

For the Rössler flow the contracting eigenvalues turn out to be insanely contracting, a factor of e^{-32} per one par-course of the attractor, so their numerical determination is quite difficult. Fortunately, they are irrelevant; for all practical purposes the strange attractor of the Rössler flow is 1-dimensional, a very good realization of a horseshoe template.

Much of this example is also worked out in Dong [1].

(G. Simon and P. Cvitanović)

[return to p. 139](#)

Example 7.2. A hyperplane Poincaré section. Let us for the sake of simplicity assume that the Poincaré section is a (hyper)-plane, i.e., it is given by the linear condition (3.5)

$$(x - x_0) \cdot \hat{n} = 0, \quad (7.10)$$

where $\hat{n}$ is a vector normal to the Poincaré section and x_0 is any point in the Poincaré section. The Newton setup is then (derived as (7.7))

$$\begin{pmatrix} 1 - J & -v(x) \\ \hat{n} & 0 \end{pmatrix} \begin{pmatrix} x' - x \\ \Delta t \end{pmatrix} = \begin{pmatrix} -F(x) \\ 0 \end{pmatrix}. \quad (7.11)$$

The last row in this equation ensures that x will be on the section surface, and the addition of $v(x)\Delta t$, a small vector along the direction of flow, ensures that such an x can be found, at least if x is sufficiently close to a fixed point of f . Alternatively, this can be solved as a least squares problem.

To illustrate that the addition of this extra constraint resolves the problem of $(1 - J)$ non-invertability, we consider a simple, particular example of a 3-d flow with the $(x, y, 0)$ -plane as the Poincaré section, $a = (0, 0, 1)$. Let all trajectories cross the Poincaré section perpendicularly, so that $v = (0, 0, v_z)$, which means that the marginally stable direction is also perpendicular to the Poincaré section. Furthermore, let the unstable direction be parallel to the x -axis and the stable direction be parallel to the y -axis. The Newton setup is now

$$\begin{pmatrix} 1 - \Lambda_u & 0 & 0 & 0 \\ 0 & 1 - \Lambda_s & 0 & 0 \\ 0 & 0 & 0 & -v_z \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \delta_x \\ \delta_y \\ \delta_z \\ \delta\tau \end{pmatrix} = \begin{pmatrix} -F_x \\ -F_y \\ -F_z \\ 0 \end{pmatrix}. \quad (7.12)$$

If one considers only the upper-left $[3 \times 3]$ matrix, then this matrix is not invertible and the equation does not have a unique solution. This $[3 \times 3]$ matrix is the one we started out with, prior to adding the constraint from (7.10). However, the full $[4 \times 4]$ matrix is invertible, as $\det(\cdot) = -v_z \det(1 - M_\perp)$, where $M_\perp$ is the $[2 \times 2]$ *monodromy matrix* for a surface of section transverse to the orbit (see sect. 5.5).

(F. Christiansen)

[return to p. 140](#)

Exercises

- 7.1. **Rössler flow cycles.** (continuation of exercise 4.4) Determine all cycles for the Rössler flow (2.30), as well as their stabilities, up to 3 Poincaré section returns. Hint: study the [video](#), and use online Python code for [Homework 3](#) of the [online course](#).

Table: The Rössler flow (2.30): The itinerary p , a periodic point $x_p = (0, y_p, z_p)$ and the expanding eigenvalue Λ_p for cycles up to topological period 3.
(J. Mathiesen, G. Simon, A. Basu)

n_p	p	y_p	z_p	Λ_e
1	1	6.091768	1.299732	-2.403953
2	01	3.915804	3.692833	-3.512007
3	001	2.278281	7.416481	-2.341923
	011	2.932877	5.670806	5.344908

- 7.2. **Inverse iteration method for a Hénon repeller.**

Table: All periodic orbits up to 6 bounces for the Hamiltonian Hénon map (7.13) with $a = 6$. Listed are the cycle itinerary, its expanding eigenvalue Λ_p , and its “center of mass.” The “center of mass” is listed because it turns out that it is often a simple rational or a quadratic irrational.

p	Λ_p	$\sum x_{p,i}$
0	0.715168×10^1	-0.607625
1	-0.295285×10^1	0.274292
10	-0.989898×10^1	0.333333
100	-0.131907×10^3	-0.206011
110	0.558970×10^2	0.539345
1000	-0.104430×10^4	-0.816497
1100	0.577998×10^4	0.000000
1110	-0.103688×10^3	0.816497
10000	-0.760653×10^4	-1.426032
11000	0.444552×10^4	-0.606654
10100	0.770202×10^3	0.151375
11100	-0.710688×10^3	0.248463
11010	-0.589499×10^3	0.870695
11110	0.390994×10^3	1.095485
100000	-0.545745×10^5	-2.034134
110000	0.322221×10^5	-1.215250
101000	0.513762×10^4	-0.450662
111000	-0.478461×10^4	-0.366025
110100	-0.639400×10^4	0.333333
101100	-0.639400×10^4	0.333333
111100	0.390194×10^4	0.548583
111010	0.109491×10^4	1.151463
111110	-0.104338×10^4	1.366025

Consider the Hénon map (3.18) for the area-preserving (“Hamiltonian”) parameter value $b = -1$. The coordinates of a periodic orbit of period n_p satisfy the equation

$$x_{p,i+1} + x_{p,i-1} = 1 - ax_{p,i}^2, \quad i = 1, \dots, n_p, \quad (7.13)$$

with the periodic boundary condition $x_{p,0} = x_{p,n_p}$. Verify that the itineraries and the stabilities of the short periodic orbits for the Hénon repeller (7.13) at $a = 6$ are as listed above.

Hint: you can use any cycle-searching routine you wish, but for the complete repeller case (all binary sequences are realized), the cycles can be evaluated simply by inverse iteration, using the inverse of (7.13)

$$x''_{p,i} = S_{p,i} \sqrt{\frac{1 - x'_{p,i+1} - x'_{p,i-1}}{a}}, \quad i = 1, \dots, n_p.$$

Here $S_{p,i}$ are the signs of the corresponding periodic point coordinates, $S_{p,i} = x_{p,i}/|x_{p,i}|$. (G. Vattay)

- 7.3. **“Center of mass” puzzle.**  Why is the “center of mass,” tabulated in exercise 7.2, often a rational number?
- 7.4. **Cycle stability, helium.** Add to the helium integrator of exercise 2.11 a routine that evaluates the expanding eigenvalue for a given cycle.

Chapter 8

Hamiltonian dynamics

Conservative mechanical systems have equations of motion that are symplectic and can be expressed in Hamiltonian form. The generic properties within the class of symplectic vector fields are quite different from those within the class of all smooth vector fields: the system always has a first integral (“energy”) and a preserved volume, and equilibrium points can never be asymptotically stable in their energy level.

— John Guckenheimer

YOU MIGHT THINK that the strangeness of contracting flows, such as the Rössler flow of figure 2.6, only concerns chemists, biomedical engineers or weathercasters; physicists do Hamiltonian dynamics, right? Now, that’s full of chaos, too! The story started with Poincaré’s restricted 3-body problem, a realization that chaos rules also in general, (non)Hamiltonian flows came much later. While it is easier to visualize aperiodic dynamics when a flow is contracting onto a lower-dimensional attracting set, there are plenty of examples of chaotic flows that do preserve the full symplectic invariance of Hamiltonian dynamics.

Here we briefly review parts of deterministic dynamics that we will need later on; symplectic invariance, canonical transformations, and stability of Hamiltonian flows. The message: Euclidean distance is meaningless for symplectic flows. Instead, the distance between two states is measured by the difference of their *phase-state actions*.



If your eventual destination are applications such as chaos in energy-conserving quantum and/or mechanical systems, read this chapter, because (a) that is mechanics, and (b) one cannot do quantum theory without it. If you work in neuroscience or Navier-Stokes fluid dynamics, skip this chapter and continue by reading about billiard dynamics in chapter 9. That requires no incantations of symplectic pairs or loxodromic quartets.



fast track:
chapter 9, p. 168

8.1 Hamiltonian flows

“... to do this business right is a thing of far greater difficulty than I was aware of.”

— Sir Isaac Newton, in a letter to Edmund Halley

(P. Cvitanović and L.V. Vela-Arevalo)

An important class of flows are Hamiltonian flows, given by a Hamiltonian $H(q, p)$ together with the Hamilton's equations of motion

appendix A4
remark 2.1

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}, \quad (8.1)$$

with the $d = 2D$ phase-space coordinates x split into the configuration space coordinates and the conjugate momenta of a Hamiltonian system with D degrees of freedom (degrees of freedom):

$$x = (\mathbf{q}, \mathbf{p}), \quad \mathbf{q} = (q_1, q_2, \dots, q_D), \quad \mathbf{p} = (p_1, p_2, \dots, p_D). \quad (8.2)$$

The equations of motion (8.1) for a time-independent, D -degrees of freedom Hamiltonian can be written compactly as

$$\dot{x}_i = \omega_{ij} H_{,j}(x), \quad H_{,j}(x) = \frac{\partial}{\partial x_j} H(x), \quad (8.3)$$

where $x = (\mathbf{q}, \mathbf{p}) \in \mathcal{M}$ is a phase-space point, and the a derivative of $(\cdot)$ with respect to x_j is denoted by comma-index notation $(\cdot)_{,j}$,

$$\omega = \begin{bmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{bmatrix}, \quad (8.4)$$

is an antisymmetric $[d \times d]$ matrix, and $\mathbf{I}$ is the $[D \times D]$ unit matrix.

The energy, or the value of the time-independent Hamiltonian function at the state space point $x = (\mathbf{q}, \mathbf{p})$ is constant along the trajectory $x(t)$,

$$\begin{aligned} \frac{d}{dt} H(\mathbf{q}(t), \mathbf{p}(t)) &= \frac{\partial H}{\partial q_i} \dot{q}_i(t) + \frac{\partial H}{\partial p_i} \dot{p}_i(t) \\ &= \frac{\partial H}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i} \frac{\partial H}{\partial q_i} = 0, \end{aligned} \quad (8.5)$$

so the trajectories lie on surfaces of constant energy, or *level sets* of the Hamiltonian $\{(q, p) : H(q, p) = E\}$. For 1-dof Hamiltonian systems, this is basically the whole story.

Figure 8.1: Phase plane of the unforced, undamped Duffing oscillator. The trajectories lie on level sets of the Hamiltonian (8.30).

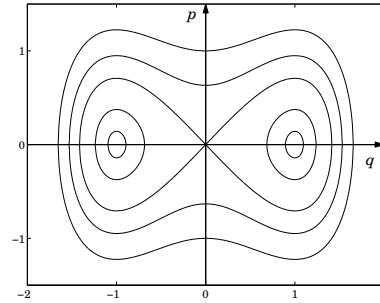
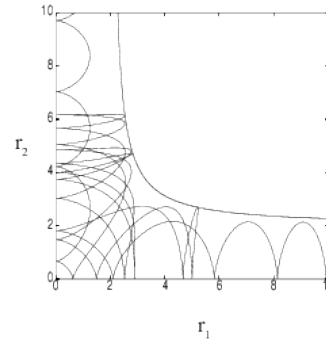


Figure 8.2: A typical collinear helium trajectory in the $[r_1, r_2]$ plane; the trajectory enters along the r_1 -axis and then, like almost every other trajectory, after a few bounces escapes to infinity, in this case along the r_2 -axis. In this example the energy is set to $H = E = -1$, and the trajectory is bounded by the kinetic energy = 0 line.



 example 8.1
p. 161

Thus all 1-dof systems are *integrable*, in the sense that the entire phase plane is stratified by curves of constant energy. These curves are either periodic, as is the case for the harmonic oscillator (a ‘bound state’), or open (a ‘scattering trajectory’). Add one more degree of freedom, and chaos breaks loose.

example A2.1

 example 8.2
p. 161

Note an important property of Hamiltonian flows: if the Hamilton equations (8.1) are rewritten in the 2D phase-space form $\dot{x}_i = v_i(x)$, the divergence of the velocity field v vanishes, namely the flow is incompressible, $\nabla \cdot v = \partial_i v_i = \omega_i H_{,ij} = 0$. The symplectic invariance requirements are actually more stringent than just the phase-space volume conservation, as we shall see in sect. 8.4.

Throughout ChaosBook we reserve the term ‘phase space’ to Hamiltonian flows. A ‘state space’ is the stage on which any flow takes place. ‘Phase space’ is a special but important case, a state space with symplectic structure, preserved by the flow. For us the distinction is necessary, as ChaosBook covers dissipative, mechanical, stochastic and quantum systems, all as one happy family.

8.2 Hamiltonian maps

Hamiltonian dynamics can be very complicated. To get across the key ideas, we need a system as simple as the Bernoulli map, but mechanical. So, we move on

from running in circles, to a mechanical rotor to kick.

8.2.1 A kicked rotor

The 1-degree of freedom maps that describe kicked rotors subject to discrete time sequences of angle-dependent force pulses $P(q_t)$, $t \in \mathbb{Z}$,



$$q_{t+1} = q_t + p_{t+1} \pmod{1}, \quad (8.6)$$

$$p_{t+1} = p_t + P(q_t), \quad (8.7)$$

with $2\pi q$ the angle of the rotor, p the momentum conjugate to the angular coordinate q , and the angular pulse $P(q_t) = P(q_{t+1}) = -V'(q_t)$ periodic with period 1. Each of these terms plays a key role in the theory of deterministic and quantum chaos in atomic and condensed matter physics, from the standard map, to the sawtooth and cat maps that we turn to now.

Eqs. (8.6, 8.7) are a discrete time form of Hamilton's equations (8.1). Eq. (8.6) is $\dot{q} = p/m$ (8.1) in terms of discrete time derivative, i.e., the configuration trajectory starting at q_t reaches $q_{t+1} = q_t + p_{t+1}\Delta t/m$ in one time step Δt . Eq. (8.7) is the time-discretized $\dot{p} = -\partial V(q)/\partial q$: at each kick the angular momentum p_t is accelerated to p_{t+1} by the force pulse $P(q_t)\Delta t$, with the time step and the rotor mass set to $\Delta t = 1$, $m = 1$.

Because values of q differing by integers are identified, whereas the corresponding values of p are not, the kicked rotor phase space is a cylinder. It is often convenient to roll it out on a plane or contract it to a torus. For the former case the “mod 1” is removed from (8.6) and for the latter it is included in (8.7).



example 8.7
p. 163



example 8.8
p. 164

8.2.2 Cat map

The simplest kicked rotor is subject to force pulses $P(q) = \mu^2 q$ proportional to the angular displacement q : in that case, the map (8.6, 8.7) is of form

$$\begin{pmatrix} q_{t+1} \\ p_{t+1} \end{pmatrix} = \mathbb{J} \begin{pmatrix} q_t \\ p_t \end{pmatrix} \pmod{1}, \quad \mathbb{J} = \begin{pmatrix} \mu^2 + 1 & 1 \\ \mu^2 & 1 \end{pmatrix}. \quad (8.8)$$

The (mod 1) makes the map a discontinuous ‘sawtooth,’ unless μ^2 is a positive integer. The map is then a Continuous Automorphism of the Torus known as the Thom-Anosov-Arnol'd-Sinai ‘cat map’, extensively studied as the simplest example of a chaotic Hamiltonian system.

remark 8.3

The determinant of the one-time-step Jacobian is $\det \mathbb{J} = 1$, i.e., the forward-in-time mapping (8.8) is area-preserving. Let $s = \text{tr } \mathbb{J} = \mu^2 + 2$ be the trace of the

Jacobian. For $|s| > 2$ the $\mathbb{J}$ characteristic equation

$$\Lambda^2 - s\Lambda + 1 = 0, \quad (8.9)$$

has real roots (Λ, Λ^{-1}) and a positive Lyapunov exponent $\lambda > 0$,

$$\Lambda = e^\lambda = \frac{1}{2}(s + \sqrt{(s-2)(s+2)}), \quad s = \text{tr } \mathbb{J} = \Lambda + \Lambda^{-1}. \quad (8.10)$$

The eigenvalues are functions of the stretching parameter s , and for $|s| > 2$ the cat map (8.8) is a fully chaotic Hamiltonian dynamical system.

A cat is Hooke's wild, 'anti-harmonic' sister. For $s < 2$ Hooke rules, with restoring oscillations around the resting state. For $s > 2$ cats rule, with exponential runaway, wrapped globally around a state space torus. Cat is to chaos what harmonic oscillator is to order. There is no more fundamental example of chaos in mechanics.



example 8.9
p. 165

8.3 Symplectic group

Either you're used to this stuff... or you have to get used to it.

—Maciej Zworski

A matrix transformation g is called *symplectic*,

$$g^\top \omega g = \omega, \quad (8.11)$$

if it preserves the *symplectic bilinear form* $\langle \hat{x}|x \rangle = \hat{x}^\top \omega x$, where $g^\top$ denotes the transpose of g , and ω is a non-singular $[2D \times 2D]$ antisymmetric matrix which satisfies

$$\omega^\top = -\omega, \quad \omega^2 = -\mathbf{1}. \quad (8.12)$$

remark 8.5

While these are defining requirements for any symplectic bilinear form, ω is often conventionally taken to be of form (8.4).



example 8.3
p. 161

If g is symplectic, so is its inverse g^{-1} . Likewise, if g_1 and g_2 are symplectic, so is their product $g_2 g_1$. Symplectic matrices form a Lie group called the *symplectic group* $\text{Sp}(d)$. Use of the symplectic group necessitates a few remarks about Lie groups in general, a topic that we study in more depth in chapter 12. A *Lie group* is a group whose elements $g(\phi)$ depend smoothly on a finite number N of parameters ϕ_a . In calculations one has to write these matrices in a specific basis, and for infinitesimal transformations they take the following form (repeated indices

are summed throughout this chapter, and the dot product refers to a sum over Lie algebra generators):

$$g(\delta\phi) \simeq 1 + \delta\phi \cdot \mathbf{T}, \quad \delta\phi \in \mathbb{R}^N, \quad |\delta\phi| \ll 1, \quad (8.13)$$

where $\{\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_N\}$, the *generators* of infinitesimal transformations, are a set of N linearly independent $[d \times d]$ matrices, which act linearly on the d -dimensional phase space $\mathcal{M}$. The infinitesimal statement of symplectic invariance follows by substituting (8.13) into (8.11) and keeping the terms linear in $\delta\phi$,

$$\mathbf{T}_a^\top \omega + \omega \mathbf{T}_a = 0. \quad (8.14)$$

This is the defining property for infinitesimal generators of *symplectic* transformations. Matrices that satisfy (8.14) are sometimes called *Hamiltonian matrices*. A linear combination of Hamiltonian matrices is a Hamiltonian matrix, so Hamiltonian matrices form a linear vector space, the *symplectic Lie algebra* $sp(d)$. By the antisymmetry of ω ,

$$(\omega \mathbf{T}_a)^\top = \omega \mathbf{T}_a. \quad (8.15)$$

is a symmetric matrix. Its number of independent elements gives the dimension (the number of independent continuous parameters) of the symplectic group $Sp(d)$,

$$N = d(d+1)/2 = D(2D+1). \quad (8.16)$$

The lowest-dimensional symplectic group $Sp(2)$, of dimension $N = 3$, is isomorphic to $SU(2)$ and $SO(3)$. The first interesting case is $Sp(3)$ whose dimension is $N = 10$.

It is easily checked that the exponential of a Hamiltonian matrix

$$g = e^{\phi \cdot \mathbf{T}} \quad (8.17)$$

is a symplectic matrix; Lie *group* elements are related to the Lie *algebra* elements by exponentiation.

question 8.4

8.4 Stability of Hamiltonian flows

Hamiltonian flows offer an illustration of the ways in which an invariance of equations of motion can affect the dynamics. In the case at hand, the *symplectic invariance* will reduce the number of independent Floquet multipliers by a factor of 2 or 4.

8.4.1 Canonical transformations

The evolution of J^t (4.5) is determined by the stability matrix A , (4.10):

$$\frac{d}{dt} J^t(x) = A(x) J^t(x), \quad A_{ij}(x) = \omega_{ik} H_{,kj}(x), \quad (8.18)$$

where the symmetric matrix of second derivatives of the Hamiltonian, $H_{,kn} = \partial_k \partial_n H$, is called the *Hessian matrix*. From (8.18) and the symmetry of $H_{,kn}$, it follows that for Hamiltonian flows (8.3)

$$A^\top \omega + \omega A = 0. \quad (8.19)$$

This is the defining property (8.14) for infinitesimal generators of *symplectic* (or canonical) transformations.

Consider now a smooth nonlinear coordinate change form $y_i = h_i(x)$ (see sect. 2.3 for a discussion), and define a ‘Kamiltonian’ function $K(x) = H(h(x))$. Under which conditions does K generate a Hamiltonian flow? In what follows we will use the notation $\partial_{\tilde{j}} = \partial/\partial y_j$, $s_{i,j} = \partial h_i/\partial x_j$. By employing the chain rule we have that

$$K_{,j} = H_{,\tilde{l}} s_{\tilde{l},j} \quad (8.20)$$

(Here, as elsewhere in this book, a repeated index implies summation.) By virtue of (8.1), $\tilde{\partial}_{\tilde{l}} H = -\omega_{lm} \dot{y}_m$, so that, again by employing the chain rule, we obtain

$$\omega_{ij} \partial_j K = -\omega_{ij} s_{j,\tilde{l}} \omega_{lm} s_{m,n} \dot{x}_n \quad (8.21)$$

The right hand side simplifies to $\dot{x}_i$ (yielding Hamiltonian structure) only if

$$-\omega_{ij} s_{\tilde{l},j} \omega_{lm} s_{m,n} = \delta_{in} \quad (8.22)$$

or, in compact notation,

$$-\omega(\partial h)^\top \omega(\partial h) = \mathbf{1} \quad (8.23)$$

which is equivalent to the requirement (8.11) that ∂h is symplectic. The transformation h is then called a *canonical transformation*. We care about canonical transformations for two reasons. First (and this is a dark art), if the canonical transformation h is very cleverly chosen, the flow in new coordinates might be considerably simpler than the original flow. Second, Hamiltonian flows themselves are a prime example of canonical transformations.

question 8.2

example A2.1



example 8.4

p. 162

8.4.2 Stability of equilibria of Hamiltonian flows

For an equilibrium point x_q , the stability matrix A is constant. Its eigenvalues describe the linear stability of the equilibrium point. A is the matrix (8.19) with real matrix elements, so its eigenvalues (the Floquet exponents of (5.1)) are either real or come in complex pairs. In the case of Hamiltonian flows, it follows from (8.19) that the characteristic polynomial of A for an equilibrium x_q satisfies

$$\begin{aligned} \det(A - \lambda \mathbf{1}) &= \det(\omega^{-1}(A - \lambda \mathbf{1})\omega) = \det(-\omega A \omega - \lambda \mathbf{1}) \\ &= \det(A^\top + \lambda \mathbf{1}) = \det(A + \lambda \mathbf{1}). \end{aligned} \quad (8.24)$$

section 5.1

exercise 8.4

exercise 8.5

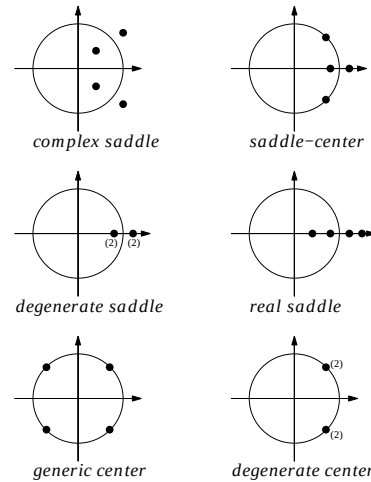


Figure 8.3: Stability of a symplectic map in $\mathbb{R}^4$.

In other words, symplectic invariance implies that $-\lambda$, λ^* and $-\lambda^*$ are also eigenvalues of the system, when λ is an eigenvalue. Distinct symmetry classes of the Floquet exponents of an equilibrium point in a 2-dof system are displayed in figure 8.4. It is worth noting that while the linear stability of equilibria in a Hamiltonian system always respects this symmetry, the nonlinear stability can be completely different.

8.5 Symplectic maps

So far we have considered only continuous time Hamiltonian flows. For finite time evolution mappings, as discussed in sect. 4.4 and sect. 4.5, the stability of maps is characterized by eigenvalues of their Jacobian matrices, or ‘multipliers.’ A multiplier $\Lambda = \Lambda(x_0, t)$ associated with a trajectory is an eigenvalue of the Jacobian matrix J . If J is symplectic, (8.11) implies that

$$J^{-1} = -\omega J^T \omega, \quad (8.25)$$

so the characteristic polynomial is reflexive. Thus, it satisfies

$$\begin{aligned} \det(J - \Lambda \mathbf{1}) &= \det(J^T - \Lambda \mathbf{1}) = \det(-\omega J^T \omega - \Lambda \mathbf{1}) \\ &= \det(J^{-1} - \Lambda \mathbf{1}) = \det(J^{-1}) \det(\mathbf{1} - \Lambda J) \\ &= \Lambda^{2D} \det(J - \Lambda^{-1} \mathbf{1}). \end{aligned} \quad (8.26)$$

Hence if Λ is an eigenvalue of J , so are $1/\Lambda$, Λ^* and $1/\Lambda^*$. Real eigenvalues always come paired as Λ , $1/\Lambda$. The Liouville conservation theorem of phase-space volumes (8.35) is an immediate consequence of this pairing of eigenvalues. The complex eigenvalues come in pairs Λ , Λ^* , $|\Lambda| = 1$, or in loxodromic quartets Λ , $1/\Lambda$, Λ^* and $1/\Lambda^*$. These possibilities are illustrated in figure 8.3.



example 8.6
p. 162



example 8.5
p. 162

8.6 Poincaré invariants

Let C be a region in phase space and $V(0)$ its volume. Denoting the flow of the Hamiltonian system by $f^t(x)$, the volume of C after a time t is $V(t) = f^t(C)$, and using (8.35) we derive the *Liouville theorem*:

$$\begin{aligned} V(t) &= \int_{f^t(C)} dx = \int_C \left| \det \frac{\partial f^t(x')}{\partial x} \right| dx' \\ &= \int_C \det(J) dx' = \int_C dx' = V(0), \end{aligned} \quad (8.27)$$

Here we see that Hamiltonian flows preserve phase-space volumes.

The symplectic structure of Hamilton's equations buys us more than 'incompressibility', or the conservation of phase-space volume. Consider the symplectic product of two infinitesimal vectors

$$\begin{aligned} \langle \delta x | \delta \hat{x} \rangle &= \delta x^\top \omega \delta \hat{x} = \delta p_i \delta \hat{q}_i - \delta q_i \delta \hat{p}_i \\ &= \sum_{i=1}^D \{ \text{oriented area in the } (q_i, p_i) \text{ plane} \}. \end{aligned} \quad (8.28)$$

Time t later we have

$$\langle \delta x' | \delta \hat{x}' \rangle = \delta x^\top J^\top \omega J \delta \hat{x} = \delta x^\top \omega \delta \hat{x}.$$

This has the following geometrical meaning. Imagine that there is a reference phase-space point. Take two other points infinitesimally close, with the vectors δx and $\delta \hat{x}$ describing their displacements relative to the reference point. Under these dynamics, the three points are mapped to three new points which are still infinitesimally close to one another. The meaning of the above expression is that the area of the parallelepiped spanned by the three final points is the same as that spanned by the initial points. The integral (Stokes theorem) version of this infinitesimal area invariance states that for Hamiltonian flows the sum of D oriented areas $\mathcal{V}_i$ bounded by D loops $\Omega \mathcal{V}_i$, one per each (q_i, p_i) plane, is conserved:

$$\int_{\mathcal{V}} dp \wedge dq = \oint_{\Omega \mathcal{V}} p \cdot dq = \text{invariant}. \quad (8.29)$$

One can show that also the $4, 6, \dots, 2D$ phase-space volumes are preserved. The phase space is $2D$ -dimensional, but as there are D coordinate combinations conserved by the flow, Hamiltonian flow is D -dimensional. Hence for Hamiltonian flows, the key notion of dimensionality is the number of the degrees of freedom—rather than phase-space dimensionality $d = 2D$.

question 8.1

question 8.3



in depth:
appendix A8.1, p. 947

Résumé

Physicists do Lagrangians and Hamiltonians. Many know of no world other than the perfect world of quantum mechanics and quantum field theory in which the energy and much else is conserved. From a dynamical point of view, Hamiltonian flow is just a flow with symmetry. The stability matrix $A_{ij} = \omega_{ik} H_{,kj}(x)$ of a Hamiltonian flow $\dot{x}_i = \omega_{ij} H_{,j}(x)$ satisfies $A^\top \omega + \omega A = 0$. Its integral along the trajectory is also the linearization of the flow J that we call the ‘Jacobian matrix’, which is symplectic. Thus, a Hamiltonian flow is a canonical transformation in the sense that the Hamiltonian time evolution $x' = f^t(x)$ is a transformation whose linearization (Jacobian matrix) $J = \partial x' / \partial x$ preserves the symplectic form, $J^\top \omega J = \omega$. This implies that A is in the symplectic algebra $sp(2D)$ and that the $2D$ -dimensional Hamiltonian phase-space flow preserves D oriented infinitesimal volumes, or Poincaré invariants. Hence, the Liouville theorem, which states that phase-space volume is conserved, is one consequence of this invariance.

Symplectic invariance enforces $|\Lambda| = 1$ for complex eigenvalue pairs and precludes the existence of attracting equilibria and limit cycles typical of dissipative flows. However, for hyperbolic equilibria and periodic orbits $|\Lambda| > 1$, and this pairing requirement only enforces a particular value on the $1/\Lambda$ contracting direction. Hence the description of chaotic dynamics as a sequence of saddle visitations is the same for the Hamiltonian and dissipative systems. You might find symplecticity beautiful. Once you understand that every time you identify symmetry means using it, you might curse the day [24] you learned to say ‘symplectic’.

chapter 12

Commentary

In theory there is no difference between theory and practice.
In practice there is.

—Anonymous

Remark 8.1. Hamiltonian dynamics, sources. If you are reading this book, in theory you already know everything that is in this chapter. In practice you do not. Try this: Put your right hand on your heart and say: “I understand why nature prefers symplectic geometry.” Honest?

Where does the skew-symmetric ω come from? Newton $f = ma$ law for a motion in a potential is $m\ddot{q} = -\partial V$. Rewrite this as a pair of first order ODEs, $\dot{q} = p/m$, $\dot{p} = -\partial V$, define the total energy $H(q, p) = p^2/2m + V(q)$, and voila, the equations of motion take on the symplectic form (8.3). What makes this important is the fact that the evolution in time (and more generally any canonical transformation) preserves this symplectic structure, as shown in sect. 8.4.1. Another way to put it: a gradient flow $\dot{x} = -\partial V(x)$ contracts a state space volume into a fixed point. When that happens, $V(x)$ is a ‘Lyapunov function’, and the equilibrium $x = 0$ is ‘Lyapunov asymptotically stable’. In contrast, the ‘-’ sign in the symplectic action on (q, p) coordinates, $\dot{p} = -\partial V$ induces a rotation and conserves phase-space areas. For a symplectic flow, there can be no volume contraction.

Out there, there are centuries of accumulated literature on Hamilton, Lagrange, Euler, Jacobi etc. formulation of mechanics, some of it excellent. In context of what we

will need here, we make a very subjective recommendation—we enjoyed reading Percival and Richards [30], Ozorio de Almeida [28], and Lowenstein [23]. Exposition of sect. 8.3 follows Dragt [9]. There are two conventions in literature for what the integer argument of $\mathrm{Sp}(\cdots)$ stands for: either $\mathrm{Sp}(D)$ or $\mathrm{Sp}(d)$ (used, for example, in refs. [7, 9]), where D = number of degrees of freedom, and $d = 2D$ dimension of the phase space. As explained in Chapter 13 of ref. [7], symplectic groups are the ‘negative dimensional’, $d \rightarrow -d$ sisters of the orthogonal groups, so only the second notation makes sense in the grander scheme of things. Mathematicians can even make sense of the d = odd-dimensional case, see Proctor [14, 31], by dropping the requirement that ω is non-degenerate, and defining a symplectic group $\mathrm{Sp}(\mathcal{M}, \omega)$ acting on a vector space $\mathcal{M}$ as a subgroup of $GL(\mathcal{M})$ which preserves a skew-symmetric bilinear form ω of *maximal possible rank*. The odd symplectic groups $\mathrm{Sp}(2D+1)$ are not semisimple. If you care about group theory for its own sake, chapter 14 of ref. [7] is fun, too. The dynamical systems symmetry reduction techniques of chapter 12 are still too primitive to be applicable to Quantum Field Theory.

Referring to the $\mathrm{Sp}(d)$ Lie algebra elements as ‘Hamiltonian matrices’, as one sometimes does [9, 38], conflicts with what is meant by a ‘Hamiltonian matrix’ in quantum mechanics. In this context, the quantum Hamiltonian is sandwiched between vectors taken from any complete set of quantum states. We are not sure where this name comes from; Dragt cites refs. [12, 15], and chapter 17 of his own book [10]. Fulton and Harris [12] used this terminology, as well as Van Loan [29] in 1981 and Tausky [37] in 1972. The terminology might go all the way back to Sylvester’s matrix theorem.

Remark 8.2. Standard map. Area-preserving maps may be used to describe kicked rotors that are subject to a discrete time sequence of angle-dependent impulses (8.48). These maps play an important role in the theory of chaos in Hamiltonian systems, from the Taylor, Chirikov and Greene standard map [4, 22] to the cat maps. Standard maps model free rotors under the influence of short periodic pulses, as can be physically implemented, for instance, by pulsed optical lattices in cold atoms physics. On the theoretical side, standard maps illustrate a number of important features: small k values provide an example of *KAM* perturbative regime (see ref. [19]), while larger k ’s illustrate deterministic chaotic transport [4, 25], and the transition to global chaos presents remarkable universality features [17, 18, 32]. The quantum counterpart of this model has been widely investigated, as the first example where phenomena like quantum dynamical localization have been observed [2]. Stability residue was introduced by Greene [17]. For some hands-on experience of the standard map, download Meiss simulation code [26].

Remark 8.3. Cat map. There is a whole strain of mathematical literature that refers to the cat map as the ‘Thom’s toral automorphism’, or ‘Thom-Anosov diffeomorphism’. We refer here to (8.52), with $s = 3$ the least unstable of the cat maps, as the “Arnol’d”, or “Thom-Arnol’d-Sinai cat map” [1, 8], and to general maps with integer $s \geq 3$ as “cat maps”. Anosov maps, and in particular cat maps, have been extensively analyzed as examples of genuine Hamiltonian chaotic evolution. Cat maps exhibit ergodicity, mixing, exponential sensitivity to variation of the initial conditions (the positivity of the Lyapunov exponent), and the positivity of the Kolmogorov-Sinai entropy [36]. Jim Crutchfield [6] online [course lect. 13](#) illustrates nicely the action of Thom-Arnol’d cat map. The course is thoughtfully put together, with notes and videos, a valuable resource. Detailed understanding of dynamics of cat maps is important also for the much richer world of nonlinear hyperbolic toral automorphisms, see refs. [5, 13, 34] for examples. (continued in remark 14.5)

example 8.8

Remark 8.4. Symplectic, canonical. The term symplectic—Greek for twining or plait-

ing together— was introduced into mathematics by Hermann Weyl. ‘Canonical’ lineage is church-doctrinal: Greek ‘kanon’, referring to a reed used for measurement, came to mean in Latin a rule or a standard. In a fine overview, Gotay and Isenberg [16] go as far as to claim that all of science will be symplectized, with continuing mathematicians’ deep dive into symplectic geometry. They might well be right.

Remark 8.5. The sign convention of ω . The overall sign of ω , the symplectic invariant in (8.3), is set by the convention that Hamilton’s principal function (for energy conserving flows) is given by $R(q, q', t) = \int_q^{q'} p_i dq_i - Et$. With this sign convention, the action along a deterministic path is minimal, and the kinetic energy of a free particle is positive. Any finite-dimensional symplectic vector space has a *Darboux basis* such that ω takes form (8.11). Dragt [9] convention for phase-space variables is as in (8.2). He calls the dynamical trajectory $x_0 \rightarrow x(x_0, t)$ the ‘transfer map’, something that we will avoid here, as it conflicts with the well established use of ‘transfer matrices’ in statistical mechanics.

Remark 8.6. Loxodromic quartets. For symplectic flows, real eigenvalues always come paired as $\Lambda, 1/\Lambda$, and complex eigenvalues come either in Λ, Λ^* pairs, $|\Lambda| = 1$, or $\Lambda, 1/\Lambda, \Lambda^*, 1/\Lambda^*$ loxodromic quartets. As most maps studied in introductory nonlinear dynamics are $2d$, you have perhaps never seen a loxodromic quartet. How likely are we to run into such things in higher dimensions? According to a very extensive study of periodic orbits of a driven billiard with a four dimensional phase space, carried in ref. [21], the three kinds of eigenvalues occur with about the same likelihood.

Question 8.1. Dream student Alice wants to know

Q Dynamics equals a Hamiltonian plus a bracket. Why don’t you just say it?

A It is true that in the tunnel vision of atomic mechanics the world is Hamiltonian. But it is more wondrous than that. This chapter starts with Newton 1687: force equals acceleration, and we always replace a higher order time derivative with a set of first order equations. If there are constraints, or fully relativistic Quantum Field Theory is your thing, the tool of choice is to recast Newton equations as a Lagrangian 1788 variational principle. If you still live in material but non-relativistic world and have not gotten beyond Heisenberg 1925, you will find Hamilton’s 1827 principal function handy. The question is not whether the world is Hamiltonian - it is not - but why is it so often profitably formulated this way. For Maupertuis 1744 variational principle was a proof of God’s existence; for Lagrange who made it mathematics, it was just a trick. Our sect. 38.1.1 “Semiclassical evolution” is an attempt to get inside 17 year old Hamilton’s head, but it is quite certain that he did not get to it the way we think about it today. He got to the ‘Hamiltonian’ by studying optics, where the symplectic structure emerges as the leading WKB approximation to wave optics; higher order corrections destroy it again. In dynamical systems theory, the densities of trajectories are transported by Liouville evolution operators, as explained here in sect. 19.6. Evolution in time is a one-parameter Lie group, and Lie groups act on functions infinitesimally by derivatives. If the evolution preserves additional symmetries, these derivatives have to respect them, and so ‘brackets’ emerge as a statement of symplectic invariance of the flow. Dynamics with a symplectic structure are just a special case of how dynamics moves densities of trajectories around. Newton is deep, Poisson brackets are technology and thus they appear naturally only by the time we get to chapter 19. Any narrative is of necessity linear, and putting Poisson ahead of Newton [35] would be a disservice to you, the student. But if you insist: Dragt and Habib [9, 11] offer a concise discussion of symplectic Lie operators and their relation to Poisson brackets.

Question 8.2. Alice, frustrated

Q I hate these $s_{m,n}$ in (8.22). Can't you use a more sensible notation?

A Be my guest.

Question 8.3. Alice, frustrated

Q Would it kill you to draw some figures in this chapter? It is supposed to be all about geometry?

A OK, isn't figure 8.5 pretty?

Question 8.4. Dream student Alice

Q Something is amiss here... The group orbit of $x \in \mathcal{M}$ is embedded into $\mathcal{M}$, so it cannot be of a higher dimension than d , but the dimension of the tangent space of the most general action of the group is $N \propto d^2$ (I'm thinking of $U(d)$, $SO(d)$ and $Sp(d)$ now), so I cannot fit all of it in a d -dimensional phase space. What gives?

A

Remark 8.7. Diagnosing chaos. In sect. 1.3.1 we have stated that a deterministic system exhibits 'chaos' if its orbits are locally unstable (positive Lyapunov exponent) and globally mixing (positive entropy). In sect. 6.2, we define Lyapunov exponents and discuss their evaluation, but already at this point it would be handy to have a few quick numerical methods to diagnose chaotic dynamics. Laskar's *frequency analysis* method [20] is useful for extracting quasi-periodic and weakly chaotic regions of state space in Hamiltonian dynamics with many degrees of freedom. For pointers to other numerical methods, see ref. [33].

References

- [1] V. I. Arnol'd and A. Avez, *Ergodic Problems of Classical Mechanics* (Addison-Wesley, Redwood City, 1989).
- [2] G. Casati and B. V. Chirikov, *Quantum Chaos: Between Order and Disorder* (Cambridge Univ. Press, Cambridge UK, 1995).
- [3] T. M. Cherry, "Some examples of trajectories defined by differential equations of a generalised dynamical type", *Trans. Camb. Phil. Soc.* **23**, 169–200 (1925).
- [4] B. V. Chirikov, "A universal instability of many-dimensional oscillator system", *Phys. Rep.* **52**, 263–379 (1979).
- [5] S. C. Creagh, "Quantum zeta function for perturbed cat maps", *Chaos* **5**, 477–493 (1995).
- [6] J. Crutchfield, *Roadmap for Natural Computation and Self-Organization*, tech. rep., Physics 256A course (U. California, Davis, 2017).
- [7] P. Cvitanović, *Group Theory: Birdtracks, Lie's and Exceptional Groups* (Princeton Univ. Press, Princeton NJ, 2008).
- [8] R. L. Devaney, *An Introduction to Chaotic Dynamical systems*, 2nd ed. (Westview Press, 2008).
- [9] A. J. Dragt, "The symplectic group and classical mechanics", *Ann. New York Acad. Sci.* **1045**, 291–307 (2005).

- [10] A. J. Dragt, *Lie Methods for Nonlinear Dynamics with Applications to Accelerator Physics*, 2011.
- [11] A. J. Dragt and S. Habib, *How Wigner functions transform under symplectic maps*, 1998.
- [12] W. Fulton and J. Harris, *Representation Theory* (Springer, New York, 1991).
- [13] I. García-Mata and M. Saraceno, “Spectral properties and classical decays in quantum open systems”, *Phys. Rev. E* **69**, 056211 (2004).
- [14] I. M. Gel’fand and A. V. Zelevinskii, “Models of representations of classical groups and their hidden symmetries”, *Funct. Anal. Appl.* **18**, 183–198 (1984).
- [15] H. Georgi, *Lie Algebras in Particle Physics* (Perseus Books, Reading, MA, 1999).
- [16] M. J. Gotay and G. A. Isenberg, “The symplectization of science”, *Gazette des Mathématiciens* **54**, 59–79 (1992).
- [17] J. M. Greene, “Two-dimensional measure-preserving mappings”, *J. Math. Phys.* **9**, 760 (1968).
- [18] J. M. Greene, “A method for determining a stochastic transition”, *J. Math. Phys.* **20**, 1183–1201 (1979).
- [19] J. V. José and E. J. Salatan, *Classical dynamics - A contemporary approach* (Cambridge Univ. Press, Cambridge UK, 1998).
- [20] J. Laskar, “The chaotic behavior of the solar system: a numerical estimate of the chaotic zones”, *Icarus* **88**, 266–291 (1990).
- [21] F. Lenz, C. Petri, F. N. R. Koch, F. K. Diakonov, and P. Schmelcher, “Evolutionary phase space in driven elliptical billiards”, *New J. Phys.* **11**, 083035 (2009).
- [22] A. J. Lichtenberg and M. A. Lieberman, *Regular and Chaotic Dynamics*, 2nd ed. (Springer, New York, 2013).
- [23] J. H. Lowenstein, *Essentials of Hamiltonian Dynamics* (Cambridge Univ. Press, Cambridge UK, 2012).
- [24] J. E. Marsden and T. S. Ratiu, *Introduction to Mechanics and Symmetry* (Springer, New York, 1999).
- [25] J. D. Meiss, “Symplectic maps, variational principles, and transport”, *Rev. Mod. Phys.* **64**, 795–848 (1992).
- [26] J. D. Meiss, “Visual explorations of dynamics: The standard map”, *Pramana, Indian Acad. Sci.* **70**, 965–988 (2008).
- [27] K. R. Meyer, “Counter-examples in dynamical systems via normal form theory”, *SIAM Rev.* **28**, 41–51 (1986).
- [28] A. M. Ozorio de Almeida, *Hamiltonian Systems: Chaos and Quantization* (Cambridge Univ. Press, Cambridge UK, 1989).
- [29] C. C. Paige and C. Van Loan, “A Schur decomposition for Hamiltonian matrices”, *Linear Algebra Appl.* **41**, 11–32 (1981).

- [30] I. Percival and D. Richards, *Introduction to Dynamics* (Cambridge Univ. Press, Cambridge, 1982).
- [31] R. A. Proctor, “Odd symplectic groups”, *Inv. Math.* **92**, 307–332 (1988).
- [32] S. J. Shenker and L. P. Kadanoff, “Critical behavior of a KAM surface: I. Empirical results”, *J. Stat. Phys.* **27**, 631–656 (1982).
- [33] C. Skokos, “Alignment indices: a new, simple method for determining the ordered or chaotic nature of orbits”, *J. Phys. A* **34**, 10029–10043 (2001).
- [34] J. Slipantschuk, O. F. Bandtlow, and W. Just, “Complete spectral data for analytic Anosov maps of the torus”, *Nonlinearity* **30**, 2667 (2017).
- [35] M. Stone and P. Goldbart, *Mathematics for Physics: A Guided Tour for Graduate Students* (Cambridge Univ. Press, Cambridge UK, 2009).
- [36] R. Sturman, J. M. Ottino, and S. Wiggins, *The Mathematical Foundations of Mixing* (Cambridge Univ. Press, 2006).
- [37] O. Taussky, “The role of symmetric matrices in the study of general matrices”, *Linear Algebra Appl.* **5**, 147–154 (1972).
- [38] Wikipedia contributors, [Hamiltonian matrix](#) — Wikipedia, The Free Encyclopedia, 2020.

8.7 Examples

Example 8.1. Unforced, undamped Duffing oscillator. When the damping term is removed from the Duffing oscillator (2.24), the system can be written in Hamiltonian form,

$$H(q, p) = \frac{p^2}{2} - \frac{q^2}{2} + \frac{q^4}{4}. \quad (8.30)$$

This is a 1-dof Hamiltonian system, with a 2-dimensional state space, the plane (q, p) . The Hamilton's equations (8.1) are

$$\dot{q} = p, \quad \dot{p} = q - q^3. \quad (8.31)$$

For 1-dof systems, the 'surfaces' of constant energy (8.5) are curves that stratify the phase plane (q, p) , and the dynamics is very simple: the curves of constant energy *are* the trajectories, as shown in figure 8.1.

[return to p. 148](#)

Example 8.2. Collinear helium. In chapter 42 we shall apply the periodic orbit theory to the quantization of helium. In particular, we will study *collinear helium*, a doubly charged nucleus with two electrons arranged on a line, an electron on each side of the nucleus. The Hamiltonian for this system is

$$H = \frac{1}{2}p_1^2 + \frac{1}{2}p_2^2 - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_1 + r_2}. \quad (8.32)$$

[chapter 42](#)

Collinear helium has 2 degrees of freedom, and thus a 4-dimensional phase space $\mathcal{M}$, which is energy conservation stratified by 3-dimensional constant energy hypersurfaces. In order to visualize it, we often project the dynamics onto the 2-dimensional configuration plane, the (r_1, r_2) , $r_i \geq 0$ quadrant, figure 8.2. It looks messy, and, indeed, it will turn out to be no less chaotic than a pinball bouncing between three disks. As always, a Poincaré section will be more informative than this rather arbitrary projection of the flow. The difference is that in such *projection* we see the flow from an arbitrary perspective, with trajectories crisscrossing. In a *Poincaré section* the flow is decomposed into intrinsic coordinates, a pair along the marginal stability time and energy directions, and the rest transverse, revealing the phase-space structure of the flow.

[return to p. 148](#)

Example 8.3. Symplectic form for $D = 2$. For two degrees of freedom, the phase space is 4-dimensional, $x = (q_1, q_2, p_1, p_2)$, and the symplectic 2-form is

$$\omega = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}. \quad (8.33)$$

The symplectic bilinear form $\langle x^{(1)} | x^{(2)} \rangle$ is the sum over the areas of the parallelepipeds spanned pairwise by components of the two vectors,

$$\langle x^{(1)} | x^{(2)} \rangle = (x^{(1)})^\top \omega x^{(2)} = (q_1^{(1)} p_1^{(2)} - q_1^{(2)} p_1^{(1)}) + (q_2^{(1)} p_2^{(2)} - q_2^{(2)} p_2^{(1)}). \quad (8.34)$$

It is this sum over oriented areas (not the Euclidean distance between the two vectors, $|x^{(2)} - x^{(1)}|$) that is preserved by the symplectic transformations.

[return to p. 150](#)

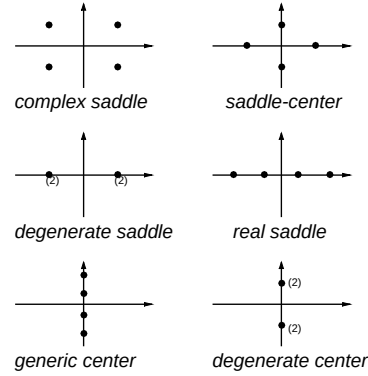


Figure 8.4: Stability exponents of a Hamiltonian equilibrium point, 2-dof.

Example 8.4. Hamiltonian flows are canonical. For Hamiltonian flows it follows from (8.19) that $\frac{d}{dt}(J^\top \omega J) = 0$, and since the initial time $J^0(x_0) = \mathbf{1}$, the Jacobian matrix is a symplectic transformation (8.11). This equality is valid for all times, so a Hamiltonian flow $f^t(x)$ is a canonical transformation, with the linearization $\partial_x f^t(x)$ a symplectic transformation (8.11). For notational brevity, we have suppressed the dependence on time and the initial point, $J = J'(x_0)$. By elementary properties of determinants, it follows from (8.11) that Hamiltonian flows are phase-space volume preserving, $|\det J| = 1$. The initial condition (4.10) for J is $J^0 = \mathbf{1}$, so one always has

$$\det J = +1. \quad (8.35)$$

[return to p. 152](#)

Example 8.5. 2-dimensional symplectic maps. In the 2-dimensional case, the eigenvalues (5.6) depend only on $\text{tr } M^t$

$$\Lambda_{1,2} = \frac{1}{2} \left(\text{tr } M^t \pm \sqrt{(\text{tr } M^t - 2)(\text{tr } M^t + 2)} \right). \quad (8.36)$$

Greene's residue criterion states that the orbit is elliptic if the *stability residue* $|\text{tr } M^t| - 2 \leq 0$, with complex eigenvalues $\Lambda_1 = e^{i\theta t}$, $\Lambda_2 = \Lambda_1^* = e^{-i\theta t}$. If $|\text{tr } M^t| - 2 > 0$ and λ is real, then the trajectory is either

$$\text{hyperbolic} \quad \Lambda_1 = e^{\lambda t}, \quad \Lambda_2 = e^{-\lambda t}, \text{ or} \quad (8.37)$$

$$\text{inverse hyperbolic} \quad \Lambda_1 = -e^{\lambda t}, \quad \Lambda_2 = -e^{-\lambda t}. \quad (8.38)$$

[return to p. 153](#)

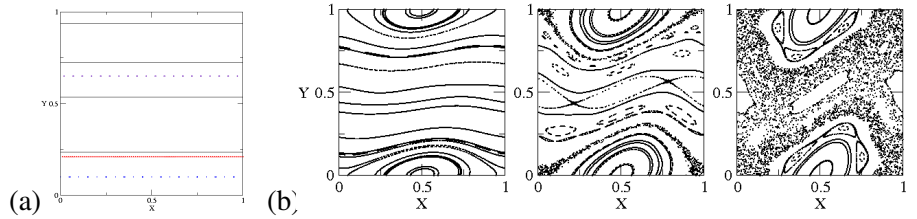
Example 8.6. Hamiltonian Hénon map. By (4.45) the Hénon map (3.18) for $b = -1$ value is the simplest 2-dimensional orientation preserving area-preserving map, often studied to better understand topology and symmetries of Poincaré sections of 2-degrees of freedom Hamiltonian flows. We find it convenient to multiply (3.19) by a and absorb the a factor into x in order to bring the Hénon map for the $b = -1$ parameter value into the form

$$x_{i+1} + x_{i-1} = a - x_i^2, \quad i = 1, \dots, n_p. \quad (8.39)$$

The 2-dimensional *Hénon map* for $b = -1$ parameter value

$$\begin{aligned} x_{n+1} &= a - x_n^2 - y_n \\ y_{n+1} &= x_n. \end{aligned} \quad (8.40)$$

Figure 8.5: Phase portrait for the standard map for (a) $k = 0$. Symbols denote periodic orbits, and full lines represent quasiperiodic orbits. (b) $k = 0.3$, $k = 0.85$ and $k = 1.4$. Each plot consists of 20 random initial conditions, each iterated 400 times.



is Hamiltonian (symplectic) in the sense that it preserves area in the $[x, y]$ plane.

In the following examples of numerical calculations, we shall fix (arbitrarily) the stretching parameter value to $a = 6$, a value large enough to guarantee that all roots of $0 = f^n(x) - x$ (periodic points) are real.

Example 8.7. Standard map. Given a smooth function $g(x)$, the map

$$\begin{aligned} x_{n+1} &= x_n + y_{n+1} \\ y_{n+1} &= y_n + g(x_n) \end{aligned} \quad (8.41)$$

is an area-preserving map. The corresponding n th iterate Jacobian matrix (4.21) is

$$M^n(x_0, y_0) = \prod_{k=n}^1 \begin{bmatrix} 1 + g'(x_k) & 1 \\ g'(x_k) & 1 \end{bmatrix}. \quad (8.42)$$

The map preserves areas, $\det M = 1$, and one can easily check that M is symplectic. In particular, one can consider x on the unit circle, and y as the conjugate angular momentum, with a function g periodic with period 1. The phase space of the map is thus the cylinder $S_1 \times \mathbf{R}$ (S_1 stands for the 1-torus, which is a fancy way to say “circle”). By taking (8.41) mod 1, the map can be reduced on the 2-torus S_2 .

The *standard map* corresponds to the choice

$$g(x) = k/2\pi \sin(2\pi x). \quad (8.43)$$

When $k = 0$, $y_{n+1} = y_n = y_0$, so that angular momentum is conserved, and the angle x rotates with uniform velocity

$$x_{n+1} = x_n + y_0 = x_0 + (n+1)y_0 \mod 1.$$

The choice of y_0 determines the nature of the motion (in the sense of sect. 2.1.1). For $y_0 = 0$, we have that every point on the $y_0 = 0$ line is stationary; for $y_0 = p/q$, the motion is periodic; and for irrational y_0 , any choice of x_0 leads to a quasiperiodic motion (see figure 8.5(a)).

Despite the simple structure of the standard map, a complete description of its dynamics for arbitrary values of the nonlinear parameter k is fairly complex. This can be appreciated by looking at phase portraits of the map for different k values. When k is very small, the phase space looks very much like a slightly distorted version of figure 8.5(a); whereas, when k is sufficiently large, single trajectories wander erratically on a large fraction of the phase space, as in figure 8.5(b).

This gives a glimpse of the typical scenario of transition to chaos for Hamiltonian systems.

exercise 9.7
return to p. 153

remark 8.2

Note that the map (8.41) provides a stroboscopic view of the flow generated by a (time-dependent) Hamiltonian

$$H(x, y; t) = \frac{1}{2}y^2 + G(x)\delta_1(t), \quad (8.44)$$

where δ_1 denotes the periodic delta function

$$\delta_1(t) = \sum_{m=-\infty}^{\infty} \delta(t - m), \quad (8.45)$$

and

$$G'(x) = -g(x). \quad (8.46)$$

Important features of this map, including transition to *global chaos* (destruction of the last invariant torus), may be tackled by detailed investigation of the stability of periodic orbits. A family of periodic orbits of period Q already present in the $k = 0$ rotation maps can be labeled by its *winding number* P/Q . The *Greene residue* describes the stability of a P/Q -cycle as

$$R_{P/Q} = \frac{1}{4} (2 - \text{tr } M_{P/Q}). \quad (8.47)$$

If $R_{P/Q} \in (0, 1)$, the orbit is elliptic; for $R_{P/Q} > 1$ the orbit is hyperbolic; and for $R_{P/Q} < 0$ it is inverse hyperbolic.

For $k = 0$, all points on the $y_0 = P/Q$ line are periodic with period Q , and the winding value P/Q and marginal stability value $R_{P/Q}$ both equal zero. As soon as $k > 0$, only a $2Q$ of such orbits survive, according to Poincaré-Birkhoff theorem. Half of them are elliptic, and the other half are hyperbolic. If we further vary k in such a way that the residue of the elliptic Q -cycle goes through 1, a bifurcation takes place, and two or more periodic orbits of higher period are generated. (continued in example 8.8)

[return to p. 149](#)

Example 8.8. Kicked rotor action. (Continued from example 8.7) Area-preserving maps that describe kicked rotors subject to a discrete time sequence of angle-dependent impulses $P(x_n)$, $t \in \mathbb{Z}$,

$$x_{n+1} = x_n + p_{n+1} \quad \text{mod } 1, \quad (8.48)$$

$$p_{n+1} = p_n + P(x_n), \quad (8.49)$$

play an important role in the theory of chaos in Hamiltonian systems, from the Taylor, Chirikov and Greene standard map (8.43), to the cat map (8.51). Here $2\pi x$ is the angle of the rotor, p is the momentum conjugate to the configuration coordinate x , $P(x) = P(x+1)$ is periodic pulse with period 1, and the time step has been set to $\Delta t = 1$. Area-preserving maps that describe kicked rotors subject to a discrete time sequence of angle-dependent impulses $P(x_n)$ of form (8.48), (8.49) have a *generating function*

[remark 8.3](#)

$$L(q_n, q_{n+1}) = \frac{1}{2}(q_n - q_{n+1})^2 - V(q_n), \quad P(q) = -\frac{dV(q)}{dq}. \quad (8.50)$$

This generating function is the discrete time Lagrangian for a particle moving in potential $V(x)$. Eq. (8.48) says that in one time step Δt the configuration trajectory starting at x_n reaches $x_{n+1} = x_n + p_{n+1}\Delta t$, and (8.49) says that at each kick the angular momentum

p_n is accelerated to p_{n+1} by the force pulse $P(x_n)\Delta t$. As the values of x differing by integers are identified, and the momentum p is unbounded, the phase space is a cylinder. However, to analyse the dynamics, one can just as well compactify the state space by folding the momentum dynamics onto a circle, by adding “mod 1” to (8.49). This reduces the dynamics to a toral automorphism acting on a $(0, 1] \times (0, 1]$ square of unit area, with opposite sides identified (continued in example 8.9).

[return to p. 149](#)

Example 8.9. Cat map. (Continued from example 8.8) The simplest example of (8.48, 8.49) is a rotor subject to a force $F(x) = Kx$ linear in the displacement x . The mod 1 added to (8.49) makes this a discontinuous “sawtooth,” unless K is an integer. In that case, the map (8.48, 8.49) is a continuous automorphism of the torus, or a “cat map” [1], a linear symplectic map on the unit 2-torus state space, $(x \mapsto Ax | x \in \mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2; A \in \text{SL}(2, \mathbb{Z}))$, with coordinates $x = (x_n, p_n)$ interpreted as the angular position variable and its conjugate momentum at time instant t . Explicitly,

$$\begin{pmatrix} x_{n+1} \\ p_{n+1} \end{pmatrix} = A \begin{pmatrix} x_n \\ p_n \end{pmatrix} \mod 1, \quad A = \begin{pmatrix} a & c \\ d & b \end{pmatrix}, \quad (8.51)$$

where a, b, c, d are integers whose precise values do not matter, as long as $\det A = 1$, so that the map is symplectic (area preserving). In that case, the map is characterized by only one invariant parameter, $s = \text{tr } A$, so consider a linear, phase space (area) preserving map of a 2-torus onto itself

$$\begin{pmatrix} x_{n+1} \\ p_{n+1} \end{pmatrix} = A \begin{pmatrix} x_n \\ p_n \end{pmatrix} \mod 1, \quad A = \begin{pmatrix} s-1 & 1 \\ s-2 & 1 \end{pmatrix}, \quad (8.52)$$

where both x_n and p_n belong to the unit interval. For integer $s = \text{tr } A > 2$, the map is referred to as a cat map. It is a fully chaotic Hamiltonian dynamical system, which can be rewritten as a 2-step difference equation in (x_{n-1}, x_n) that takes a particularly simple form

[remark 8.3](#)

$$x_{n+1} - s x_n + x_{n-1} = -m_n, \quad (8.53)$$

with the unique integer “winding number” m_n at every time step n ensuring that x_{n+1} lands in the unit interval. While the dynamics is linear, the nonlinearity comes through the mod 1 operation, encoded in $m_n \in \mathcal{A}$, where $\mathcal{A}$ is finite alphabet of possible values for m_n .

A cat map is a fully chaotic Hamiltonian dynamical system, if its stability multipliers $(\Lambda^+, \Lambda^-) = (\Lambda, \Lambda^{-1})$, and

$$\Lambda^\pm = \frac{1}{2}(s \pm \sqrt{D}), \quad \Lambda = \frac{1}{2}(s + \sqrt{(s-2)(s+2)}) = e^\lambda, \quad (8.54)$$

are real. In this equation, $s = \text{tr } A = \Lambda + \Lambda^{-1}$, $\sqrt{D} = \Lambda - \Lambda^{-1}$, discriminant $D = s^2 - 4$, with a positive Lyapunov exponent $\lambda > 0$. The eigenvalues are functions of the “stretching” parameter s , and the map is chaotic if and only if $|s| > 2$. (continued in example 14.12)

[return to p. 150](#)

Exercises

- 8.1. **Complex nonlinear Schrödinger equation.** Consider the complex nonlinear Schrödinger equation in one spatial dimension [24]:

$$i \frac{\partial \phi}{\partial t} + \frac{\partial^2 \phi}{\partial x^2} + \beta \phi |\phi|^2 = 0, \quad \beta \neq 0.$$

- (a) Show that the function $\psi : \mathbb{R} \rightarrow \mathbb{C}$ defining the traveling wave solution $\phi(x, t) = \psi(x - ct)$ for $c > 0$ satisfies a second-order complex differential equation equivalent to a Hamiltonian system in $\mathbb{R}^4$ relative to the noncanonical symplectic form whose matrix is given by

$$w_c = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & -c \\ 0 & -1 & c & 0 \end{bmatrix}.$$

- (b) Analyze the equilibria of the resulting Hamiltonian system in $\mathbb{R}^4$ and determine their linear stability properties.
- (c) Let $\psi(s) = e^{ics/2} a(s)$ for a real function $a(s)$ and determine a second order equation for $a(s)$. Show that the resulting equation is Hamiltonian and has heteroclinic orbits for $\beta < 0$. Find them.
- (d) Find ‘soliton’ solutions for the complex nonlinear Schrödinger equation.

(Luz V. Vela-Arevalo)

- 8.2. **Symplectic vs. Hamiltonian matrices.** In the language of group theory, symplectic matrices form the symplectic Lie *group* $\text{Sp}(d)$, while the Hamiltonian matrices form the symplectic Lie *algebra* $\mathfrak{sp}(d)$, or the algebra of generators of infinitesimal symplectic transformations. This exercise illustrates the relation between the two:

- (a) Show that if a constant matrix A satisfy the Hamiltonian matrix condition (8.14), then $J(t) = \exp(tA)$, $t \in \mathbb{R}$, satisfies the symplectic condition (8.11), i.e., $J(t)$ is a symplectic matrix.
- (b) Show that if matrices $\mathbf{T}_a$ satisfy the Hamiltonian matrix condition (8.14), then $g(\phi) = \exp(\phi \cdot \mathbf{T})$, $\phi \in \mathbb{R}^N$, satisfies the symplectic condition (8.11), i.e., $g(\phi)$ is a symplectic matrix.

(A few hints: (i) expand $\exp(A)$, $A = \phi \cdot \mathbf{T}$, as a power series in A . Or, (ii) use the linearized evolution equation (8.18).)

- 8.3. **When is a linear transformation canonical?**

- (a) Let A be a $[n \times n]$ invertible matrix. Show that the map $\phi : \mathbb{R}^{2n} \rightarrow \mathbb{R}^{2n}$ given by $(\mathbf{q}, \mathbf{p}) \mapsto (A\mathbf{q}, (A^{-1})^\top \mathbf{p})$ is a canonical transformation.
- (b) If $\mathbf{R}$ is a rotation in $\mathbb{R}^3$, show that the map $(\mathbf{q}, \mathbf{p}) \mapsto (\mathbf{R}\mathbf{q}, \mathbf{R}\mathbf{p})$ is a canonical transformation.

(Luz V. Vela-Arevalo)

8.4. **Determinants of symplectic matrices.** Show that the determinant of a symplectic matrix is $+1$, by going through the following steps:

- (a) use (8.26) to prove that for eigenvalue pairs each member has the same multiplicity (the same holds for quartet members),
- (b) prove that the *joint* multiplicity of $\lambda = \pm 1$ is even,
- (c) show that the multiplicities of $\lambda = 1$ and $\lambda = -1$ cannot be both odd. Hint: write

$$P(\lambda) = (\lambda - 1)^{2m+1}(\lambda + 1)^{2l+1}Q(\lambda)$$

and show that $Q(1) = 0$.

8.5. **Cherry's example.** What follows refs. [3, 27] is mostly a reading exercise, about a Hamiltonian system that is *linearly stable* but *nonlinearly unstable*. Consider the Hamiltonian system on $\mathbb{R}^4$ given by

$$H = \frac{1}{2}(q_1^2 + p_1^2) - (q_2^2 + p_2^2) + \frac{1}{2}p_2(p_1^2 - q_1^2) - q_1q_2p_1.$$

- (a) Show that this system has an equilibrium at the origin, which is linearly stable. (The linearized system consists of two uncoupled oscillators with frequencies in ratios 2:1).
- (b) Convince yourself that the following is a family of solutions parameterize by a constant τ :

$$\begin{aligned} q_1 &= -\sqrt{2}\frac{\cos(t-\tau)}{t-\tau}, & q_2 &= \frac{\cos 2(t-\tau)}{t-\tau}, \\ p_1 &= \sqrt{2}\frac{\sin(t-\tau)}{t-\tau}, & p_2 &= \frac{\sin 2(t-\tau)}{t-\tau}. \end{aligned}$$

These solutions clearly blow up in a finite time; however, they start at $t = 0$ at a distance $\sqrt{3}/\tau$ from the origin, so by choosing τ large, we can find solutions starting arbitrarily close to the origin, yet going to infinity in a finite time, so the origin is *nonlinearly unstable*.

(Luz V. Vela-Arevalo)

Chapter 9

Billiards

THE DYNAMICS that we have the best intuitive grasp on, and find easiest to grapple with both numerically and conceptually, is the dynamics of billiards. For billiards, discrete time is altogether natural; a particle moving through a billiard suffers a sequence of instantaneous kicks, and executes simple motion in between, so there is no need to contrive a Poincaré section. We have already used this system in sect. 1.3 as the most intuitive and accessible example of chaos. Here we define billiard dynamics more precisely, anticipating the applications to come. As billiards lend themselves naturally to visualization, this chapter –for once– is better grasped by following the video links on the margins, then by decoding the notation by reading this text.



9.1 Billiard dynamics



A billiard is defined by a connected region $Q \subset \mathbb{R}^D$, with boundary $\partial Q \subset \mathbb{R}^{D-1}$ separating Q from its complement $\mathbb{R}^D \setminus Q$. The region Q can consist of one compact, finite volume component (in which case the billiard phase space is bounded, as for the stadium billiard of figure 9.1), or can be infinite in extent, with its complement $\mathbb{R}^D \setminus Q$ consisting of one or several finite or infinite volume components (in which case the phase space is open, as for the 3-disk pinball game in figure 1.1). In what follows, we often restrict our attention to *planar billiards*.

A point particle of mass m and momentum $p = mv$ moves freely within the billiard, along a straight line, until it encounters the boundary. There it reflects specularly (*specular* = mirrorlike), with no change in the tangential component of momentum, and instantaneous reversal of the incoming momentum p^- component normal to the boundary,

$$p = p^- - 2(p^- \cdot \hat{n}) \hat{n}, \quad (9.1)$$

with $\hat{n}$ the unit vector normal to the boundary ∂Q at the collision point. The angle of incidence equals the angle of reflection, as illustrated in figure 9.2. A billiard is

Figure 9.1: The stadium billiard is a 2-dimensional domain bounded by two semi-circles of radius $d = 1$ connected by two straight walls of length $2a$. At the points where the straight walls meet the semi-circles, the curvature of the border changes discontinuously; these are the only singular points of the flow. The length a is the only parameter.

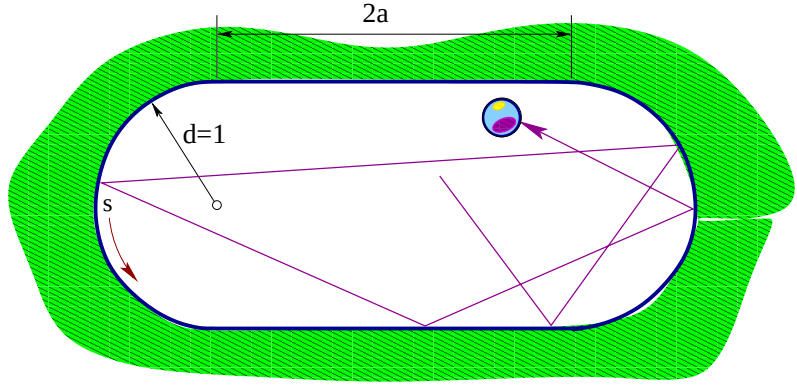
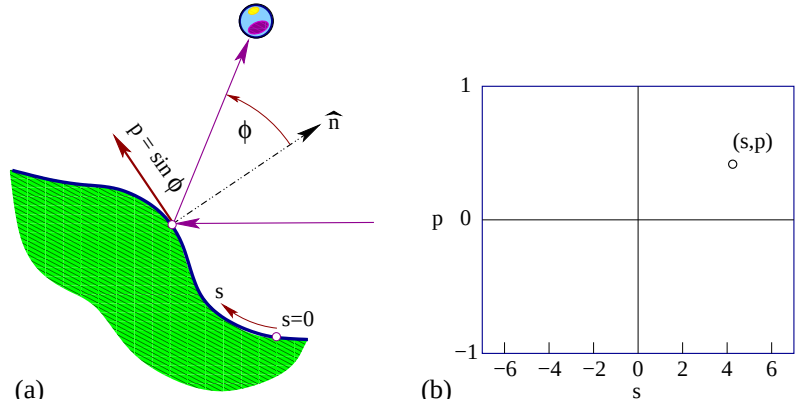


Figure 9.2: (a) A planar billiard trajectory is fixed by specifying the perimeter arclength parametrized by s and the outgoing trajectory angle ϕ , both measured counterclockwise with respect to the outward normal $\hat{n}$. (b) The Birkhoff phase-space coordinate pair (s, p) fully specifies the trajectory, where $p = |p| \sin \phi$ is the momentum component tangential to the boundary. As the pinball kinetic energy is conserved in elastic scattering, the pinball mass and the magnitude of the pinball momentum are customarily set to $m = |p| = 1$.



a Hamiltonian system with a $2D$ -dimensional phase space $x = (q, p)$ and potential $V(q) = 0$ for $q \in Q$, $V(q) = \infty$ for $q \in \partial Q$. As the energy is conserved, we can set $m = |v| = |p| = 1$ without loss of generality.



A billiard flow has a natural Poincaré section defined by Birkhoff coordinates s_n , the arc length position of the n th bounce measured along the billiard boundary, and $p_n = |p| \sin \phi_n$, the momentum component parallel to the boundary, where ϕ_n is the angle between the outgoing trajectory and the normal to the boundary. We measure both the arc length s , and the parallel momentum p counterclockwise relative to the outward normal (see figure 9.2 as well as figure 15.15 (a)). In $D = 2$, the Poincaré section is a cylinder (an annulus), figure 9.3, where the parallel momentum p ranges from $-|p|$ to $|p|$, and the s coordinate is cyclic along each connected component of ∂Q . The volume in the full phase space is preserved by the Liouville theorem (8.27). The Birkhoff coordinates $x = (s, p) \in \mathcal{P}$, are the natural choice, because with them the return map preserves the phase-space volume of the (s, p) parameterized Poincaré section (a perfectly good, often used coordinate set (s, ϕ) does not do that).

remark 2.1

exercise 9.7

exercise 9.7
section 9.2

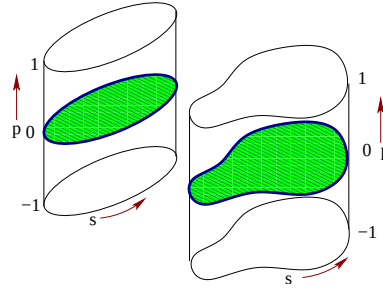
Poincaré section condition eliminates one dimension, and the energy conservation $|p| = 1$ eliminates another, so the Poincaré section return map P is $(2D - 2)$ -dimensional.



The dynamics is given by the return map

$$P : (s_n, p_n) \mapsto (s_{n+1}, p_{n+1}) \quad (9.2)$$

Figure 9.3: In $D = 2$, the billiard Poincaré section is a cylinder, where the parallel momentum p ranges over $p \in \{-1, 1\}$ and the s coordinate is cyclic along each connected component of ∂Q . The rectangular figure 9.2(b) is an unfolded cylinder, with periodic boundary conditions gluing together the left and right edge of the rectangle.



from the n th collision to the $(n + 1)$ collision. The discrete time dynamics map P is equivalent to a Hamiltonian flow (8.1), in the sense that they both describe the same full trajectory. Let t_n denote the instant of the n th collision. Then the position of the pinball $\in Q$ at time $t_n + \tau \leq t_{n+1}$ is given by $2D - 2$ Poincaré section coordinates $(s_n, p_n) \in \mathcal{P}$ together with τ , the distance reached by the pinball along the n th section of its trajectory (as we have set the pinball speed to 1, the time of flight equals the distance traversed).

 **example 9.1**
p. 175

9.2 Stability of billiards



We turn next to the question of local stability of discrete time billiard systems. Infinitesimal equations of variations (4.2) do not apply, but the multiplicative structure (4.20) of the finite-time Jacobian matrices does. As they are more physical than most maps studied by dynamicists, let us work out the billiard stability in some detail.

On the face of it, a plane billiard phase space is 4-dimensional. However, one dimension can be eliminated by energy conservation, and the other by cutting trajectories by a Poincaré section. We shall now show how going to a local frame of motion leads to a $[2 \times 2]$ Jacobian matrix.

Consider a planar billiard with phase-space coordinates $x = (q_1, q_2, p_1, p_2)$. Let t_n be the instant of the n th collision of the pinball with the billiard boundary, and $t_n^\pm = t_n \pm \epsilon$, ϵ positive and infinitesimal. With the mass and speed equal to 1, the momentum direction can be specified by angle θ : $x = (q_1, q_2, \sin \theta, \cos \theta)$. Now we parameterize the 2-dimensional neighborhood of a trajectory segment by $\delta x = (\delta z, \delta \theta)$, where

$$\delta z = \delta q_1 \cos \theta - \delta q_2 \sin \theta, \quad (9.3)$$

$\delta \theta$ is the variation in the direction of the pinball motion. Due to energy conservation, there is no need to keep track of $\delta q_{\parallel}$, variation along the flow, as that remains constant. The set of coordinates $(\delta q_1, \delta q_2)$ represents the variation in q_1 and q_2 that are transverse to the n th segment of flow. From Hamilton's equations of motion for a free particle, $dq_i/dt = p_i$, $dp_i/dt = 0$, we obtain the equations of motion

(4.1) for the linearized neighborhood

$$\frac{d}{dt}\delta\theta = 0, \quad \frac{d}{dt}\delta z = \delta\theta. \quad (9.4)$$

Let

$$(\delta z_n, \delta\theta_n) = (\delta z(t_n^+), \delta\theta(t_n^+)), \quad (\delta z_n^-, \delta\theta_n^-) = (\delta z(t_n^-), \delta\theta(t_n^-)) \quad (9.5)$$

be the local coordinates immediately after and before the n th collision, respectively. Integrating the free flight from t_{n-1}^+ to t_n^- we obtain

$$\begin{aligned} \delta z_n^- &= \delta z_{n-1} + \tau_n \delta\theta_{n-1}, & \tau_n &= t_n - t_{n-1} \\ \delta\theta_n^- &= \delta\theta_{n-1}. \end{aligned} \quad (9.6)$$

Additionally, the Jacobian matrix (4.19) for the n th free flight segment is

$$M_T(x_n) = \begin{bmatrix} 1 & \tau_n \\ 0 & 1 \end{bmatrix}. \quad (9.7)$$

To compute the reflection Jacobian matrix, think of the incoming rays as a flashlight shining on the billiard boundary at an angle; its footprint, of arclength δs , is wider than the incoming beam of width δz_n^- . At incidence angle ϕ_n (the angle between the outgoing particle and the outgoing normal to the billiard edge), the incoming transverse variation δz_n^- projects onto an arc on the billiard boundary of arclength $\delta s_n = \delta z_n^- / \cos \phi_n$. Approximating locally a smooth boundary by a circle of radius ρ_n , the angle of incidence corresponding to this arc is $\delta s_n = \rho_n \delta\phi_n$, so $\delta\phi_n = \delta z_n^- / \rho_n \cos \phi_n$. The specular law of reflection (9.1) doubles this angle and changes its orientation, increasing the angular spread of the beam to

$$\begin{aligned} \delta z_n &= -\delta z_n^- \\ \delta\theta_n &= -\delta\theta_n^- - \frac{2}{\rho_n \cos \phi_n} \delta z_n^-, \quad \rho_n = \text{local radius of curvature}, \end{aligned} \quad (9.8)$$

so the Jacobian matrix associated with the reflection is

$$M_R(x_n) = -\begin{bmatrix} 1 & 0 \\ r_n & 1 \end{bmatrix}, \quad r_n = \frac{2}{\rho_n \cos \phi_n}. \quad (9.9)$$

The full Jacobian matrix for n_p consecutive bounces describes a beam of trajectories defocused by M_T along the free flight (the τ_n terms below) and defocused/refocused at reflections by M_R (the r_n terms below)

exercise 9.6

$$M_p = (-1)^{n_p} \prod_{n=n_p}^1 \begin{bmatrix} 1 & \tau_n \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ r_n & 1 \end{bmatrix}, \quad (9.10)$$

where τ_n is the flight time of the n th free-flight segment of the orbit, $r_n = 2/\rho_n \cos \phi_n$ is the amount of defocusing due to the n th reflection, and ρ_n is the radius of curvature of the billiard boundary at the n th scattering point (for our 3-disk game

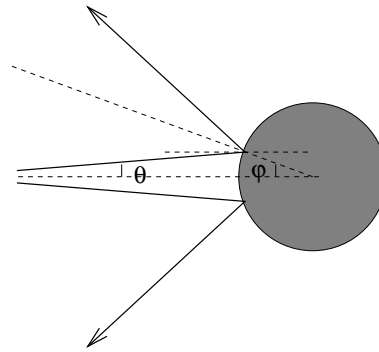


Figure 9.4: Defocusing of a beam of nearby trajectories at a billiard collision. (A. Wirzba)

of pinball, $\rho = 1$). Since billiard dynamics is phase-space volume preserving, $\det M = 1$ and the associated eigenvalues are given by (8.36).

This is an example of the Jacobian matrix chain rule (4.22) for discrete time systems, and Hénon map stability (4.44) is a second example. The stability of every flight segment or reflection taken alone is a shear with two unit eigenvalues,

$$\det M_T = \det \begin{bmatrix} 1 & \tau_n \\ 0 & 1 \end{bmatrix}, \quad \det M_R = \det \begin{bmatrix} 1 & 0 \\ r_n & 1 \end{bmatrix}, \quad (9.11)$$

but acting in concert in the interwoven sequence (9.10) they can lead to a hyperbolic deformation of the infinitesimal neighborhood of a billiard trajectory.

As a concrete application, consider the 3-disk pinball system of sect. 1.3. Analytic expressions for the periods and eigenvalues of $\bar{0}$, $\bar{1}$ and $\bar{10}$ cycles follow from elementary geometrical considerations. Longer cycles require numerical evaluation by methods such as those described in chapter 16.

exercise 16.6



exercise 16.7
exercise 9.4
chapter 16

Résumé

A particular natural application of the Poincaré section method is the reduction of a billiard flow to a boundary-to-boundary return map. The 3-disk game of pinball is to chaotic dynamics what a pendulum is to integrable systems; the simplest physical example that captures the essence of chaos. What next? For an overview of where are we now, and how are billiards going to help us, click on the video link.



Commentary

Remark 9.1. Billiards. Birkhoff coordinates [1] were introduced by -well- G.D. Birkhoff in 1927. That the 3-disk game of pinball is a quintessential example of deterministic chaos appears to have been first noted by B. Eckhardt [6]. The model was studied in depth deterministically, semiclassically and quantum mechanically by P. Gaspard and

S.A. Rice [9], and used by P. Cvitanović and B. Eckhardt [4] to demonstrate applicability of cycle expansions to quantum mechanical problems. It has been used to study the higher order $\hbar$ corrections to the Gutzwiller quantization by P. Gaspard and D. Alonso Ramirez [8], construct semiclassical evolution operators and entire spectral determinants by P. Cvitanović and G. Vattay [5], and incorporate the diffraction effects into the periodic orbit theory by G. Vattay, A. Wirzba and P.E. Rosenqvist [16]. The full quantum mechanics and semiclassics of scattering systems is developed here in the 3-disk scattering context in chapter 40. Gaspard’s monograph [7], which we warmly recommend, utilizes the 3-disk system in much more depth than will be attained here. The stadium billiard was introduced by Bunimovich [2, 3] to demonstrate that full ergodicity is possible even for billiards with focusing wall segments. Lansel, Porter and Bunimovich [10], Ranković and Porter [14], and Sawada and Taniguchi [15] discuss 2-particle billiards. For further links check ChaosBook.org/extras.

A pinball game does miss a number of important aspects of chaotic dynamics: generic bifurcations in smooth flows, the interplay between regions of stability and regions of chaos, intermittency phenomena, and the renormalization theory of the ‘border of order’ between these regions. To study these we shall have to face up to a much harder challenge, the dynamics of smooth flows. Nevertheless, pinball scattering is relevant to smooth potentials. The game of pinball may be thought of as the infinite potential wall limit of a smooth potential, and pinball symbolic dynamics can serve as a *covering* symbolic dynamics in smooth potentials. One may start with the infinite wall limit and adiabatically relax an unstable cycle onto the corresponding one for the potential under investigation. If things go well, the cycle will remain unstable and isolated, no new orbits (unaccounted for by the pinball symbolic dynamics) will be born, and the lost orbits will be accounted for by a set of pruning rules. The validity of this adiabatic approach has to be checked carefully in each application, as things can easily go wrong; for example, near a bifurcation the same naive symbol string assignments can refer to a whole island of distinct periodic orbits.



section 34.1

Another contender for the title of the ‘harmonic oscillator of chaos’ is the baker’s map which is used as the red thread through Ott’s introduction to chaotic dynamics [13]. The baker’s map is the simplest reversible dynamical system which is hyperbolic and has positive entropy. We will not have much use for the baker’s map here, as due to its piecewise linearity it is so nongeneric that it misses all of the subtleties of cycle expansion curvature corrections that will be central to this treatise.

chapter 23

Remark 9.2. Stability analysis. Chapter 1 of Gaspard’s monograph [7] is recommended reading if you are interested in Hamiltonian flows, and billiards in particular. A. Wirzba has generalized the stability analysis of sect. 9.2 to the scattering of 3-dimensional spheres (follow the links in ChaosBook.org/extras). A clear discussion of linear stability for the general d -dimensional case is given in Gaspard [7], sect. 1.4.

References

- [1] G. D. Birkhoff, “On the periodic motions of dynamical systems”, *Acta Math.* **50**, Reprinted in ref. [12], 359–379 (1927).
- [2] L. A. Bunimovich, “On ergodic properties of certain billiards”, *Funct. Anal. Appl.* **8**, 254–255 (1974).
- [3] L. A. Bunimovich, “On the ergodic properties of nowhere dispersing billiards”, *Commun. Math. Phys.* **65**, 295–312 (1979).

- [4] P. Cvitanović and B. Eckhardt, “Periodic orbit quantization of chaotic systems”, *Phys. Rev. Lett.* **63**, 823–826 (1989).
- [5] P. Cvitanović and G. Vattay, “Entire Fredholm determinants for evaluation of semi-classical and thermodynamical spectra”, *Phys. Rev. Lett.* **71**, 4138–4141 (1993).
- [6] B. Eckhardt, “Fractal properties of scattering singularities”, *J. Phys. A* **20**, 5971–5979 (1987).
- [7] P. Gaspard, *Chaos, Scattering and Statistical Mechanics* (Cambridge Univ. Press, Cambridge UK, 1998).
- [8] P. Gaspard and D. Alonso, “Ruelle classical resonances and dynamical chaos: The three- and four-disk scatterers”, *Phys. Rev. A* **45**, 8383 (1992).
- [9] P. Gaspard and S. A. Rice, “Scattering from a classically chaotic repeller”, *J. Chem. Phys.* **90**, 2225–2241 (1989).
- [10] S. Lansel, M. A. Porter, and L. A. Bunimovich, “One-particle and few-particle billiards”, *Chaos* **16**, 013129 (2006).
- [11] J. Leys, É. Ghys, and A. Alvarez, “Chapter V : Billiards - Duhem’s bull”, in *Chaos – A Mathematical Adventure* (É, Lyon, 2013).
- [12] R. S. MacKay and J. D. Meiss, *Hamiltonian Dynamical Systems: A Reprint Selection* (Taylor & Francis, Oxford, 1987).
- [13] E. Ott, *Chaos and Dynamical Systems* (Cambridge Univ. Press, Cambridge UK, 2002).
- [14] S. Ranković and M. A. Porter, “Two-particle circular billiards versus randomly perturbed one-particle circular billiards”, *Chaos* **23**, 013123 (2013).
- [15] S.-i. Sawada and T. Taniguchi, “Chaos and ergodicity of two hard disks within a circular billiard”, *Phys. Rev. E* **88**, 022907 (2013).
- [16] G. Vattay, A. Wirzba, and P. E. Rosenqvist, “Periodic orbit theory of diffraction”, *Phys. Rev. Lett.* **73**, 2304–2307 (1994).

9.3 Examples

Example 9.1. 3-disk game of pinball. In the case of bounces off a circular disk, the position coordinate $s = r\theta$ is given by angle $\theta \in [0, 2\pi]$. For example, for the 3-disk game of pinball of figure 1.6 and figure 15.15 (a) we have two types of collisions:

exercise 9.1

$$P_0 : \begin{cases} \phi' = -\phi + 2 \arcsin p \\ p' = -p + \frac{a}{R} \sin \phi' \end{cases} \quad \text{back-reflection} \quad (9.12)$$

$$P_1 : \begin{cases} \phi' = \phi - 2 \arcsin p + 2\pi/3 \\ p' = p - \frac{a}{R} \sin \phi' \end{cases} \quad \text{reflect to 3rd disk.} \quad (9.13)$$

Here a = radius of a disk, and R = center-to-center separation. Actually, as in this example we are computing intersections of circles and straight lines, nothing more than high-school geometry is required. There is no need to compute arcsin - one only needs to compute one square root per each reflection, and the simulations can be very fast.

exercise 9.2

Trajectory of the pinball in the 3-disk billiard is generated by a series of P_0 's and P_1 's. At each step one has to check whether the trajectory intersects the desired disk (and no disk in-between). With minor modifications, the above formulas are valid for any smooth billiard as long as we replace a by the local curvature of the boundary at the point of collision.

return to p. 170

Example 9.2. Duhem's bull. From Jos Leys, Étienne Ghys and Aurélien Alvarez [11] comes *Chaos – A Mathematical Adventure*, a math movie. [Chapter V Billiards - Duhem's bull](#) goes beyond the billiards discussed here, and motivates much of the symbolic dynamics to be developed below. The movie is mathematically sophisticated, and breathtakingly beautiful.

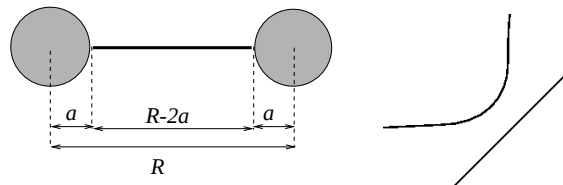


Exercises

- 9.1. **A pinball simulator.** Implement the disk $\rightarrow$ disk maps to compute a trajectory of a pinball for a given starting point, and a given $R:a$ = (center-to-center distance):(disk radius) ratio for a 3-disk system. As this requires only computation of intersections of lines and circles together with specular reflections, implementation should be within reach of a high-school student. Please start working on this program now; it will be continually expanded in chapters to come, incorporating the Jacobian calculations, Newton root-finding, and so on.

Fast code will use elementary geometry (only one $\sqrt{\cdots}$ per iteration, rest are multiplications) and eschew trigonometric functions. Provide a graphic display of the trajectories and of the Poincaré section iterates. To be able to compare with the numerical results of coming chapters, work with $R:a = 6$ and/or 2.5 values. Draw the correct versions of figure 1.8 or figure 15.3 for $R:a = 2.5$ and/or 6.

- 9.2. **Trapped orbits.** Shoot 100,000 trajectories from one of the disks, and trace out the strips of figure 1.8 for various $R:a$ by color coding the initial points in the Poincaré section by the number of bounces preceding their escape. Try also $R:a = 6:1$, though that might be too thin and require some magnification. The initial conditions can be randomly chosen, but need not - actually a clearer picture is obtained by systematic scan through regions of interest.
- 9.3. **Stability of billiard cycles.** Compute the Floquet multipliers $\{\Lambda_u, \Lambda_s\}$ for a few simple cycles:



- (a) A simple scattering billiard is the two-disk billiard. It consists of a disk of radius one centered at the origin and another disk of unit radius located at distance $L + 2$. Find all periodic orbits for this system and compute their stabilities. (You might have done this already in exercise 1.2; at least now you will be able to see where you went wrong when you knew nothing about cycles and their extraction.)
- (b) Find all periodic orbits and their stabilities for a billiard ball bouncing between the diagonal $y = x$ and one of the hyperbola branches $y = -1/x$. (continued as exercise 16.4)
- 9.4. **Pinball stability.** Add to your exercise 9.1 pinball simulator a routine that computes the $[2 \times 2]$ Jacobian matrix. To be able to compare with the numerical results of coming chapters, work with $R:a = 6$ and/or 2.5 values.
- 9.5. **A test of your pinball simulator.** Test your exercise 9.4 pinball simulator by computing numerically cycle stabilities by tracking distances to nearby orbits. Compare your result with the exact analytic formulas of exercise 16.6 and 16.7.
- 9.6. **Stadium billiard.** Consider the *Bunimovich stadium* defined in figure 9.1. The Jacobian matrix associated with the reflection is given by (9.9). Here we take $\rho_k = -1$ for the semicircle sections of the boundary, and $\cos \phi_k$ remains constant for all bounces in a rotation sequence.

The time of flight between two semicircle bounces is $\tau_k = 2 \cos \phi_k$. The Jacobian matrix of one semicircle reflection followed by the flight to the next bounce is

$$\begin{aligned} \mathbf{J} &= (-1) \begin{bmatrix} 1 & 2 \cos \phi_k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2/\cos \phi_k & 1 \end{bmatrix} \\ &= (-1) \begin{bmatrix} -3 & 2 \cos \phi_k \\ 2/\cos \phi_k & 1 \end{bmatrix}. \end{aligned}$$

A free flight must always be followed by $k = 1, 2, 3, \dots$ bounces along a semicircle, hence the natural symbolic dynamics for this problem is n -ary, with the corresponding Jacobian matrix given by shear (*ie.* the eigenvalues remain equal to 1 throughout the whole rotation), and k bounces inside a circle lead to

$$\mathbf{J}^k = (-1)^k \begin{bmatrix} -2k - 1 & 2k \cos \phi \\ 2k/\cos \phi & 2k - 1 \end{bmatrix}. \quad (9.14)$$

The Jacobian matrix of a cycle p of period n_p is given by

$$\mathbf{J}_p = (-1)^{\sum n_k} \prod_{k=1}^{n_p} \begin{bmatrix} 1 & \tau_k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ n_k r_k & 1 \end{bmatrix}. \quad (9.15)$$

Adopt your pinball simulator to the stadium billiard.

- 9.7. **Birkhoff coordinates.** Prove that the Birkhoff coordinates are phase-space volume preserving.

Chapter 10

Flips, slides and turns

A detour of a thousand pages starts with a single misstep.
—Chairman Miaw

DYNAMICAL SYSTEMS often come equipped with symmetries, such as the reflection and rotation symmetries of various potentials. This chapter assumes familiarity with basic group theory, as discussed in appendix [A10.1](#). We find the abstract notions easier to motivate by working out the examples; links to these examples are interspersed throughout the chapter. Working through them is essential to appreciate the necessity of every one of these definitions.



Question 10.1. Alice asks

Q I am not familiar with group theory. As fascinating as this is, it is akin to drinking out of a fire hose, so what should I take away from chapters [10](#) to [13](#), chapters [25](#) and [26](#) that will help me understand subsequent topics?

A In ChaosBook we invoke only those bits of group theory that are used in this book, but no matter how one tries, the definitions and the jargon pile up into a wall impenetrable to intuition. Wilkinson covers the essential ideas in a 50-minute lecture entitled [Asymmetry in dynamics](#) - do have a look. Then do what we all do - skim through chapters [10](#) to [13](#). When you start working through exercises, backtrack to find the bits that you need. Much of data science is running computers without thinking, but in order to make headway on hard problems (turbulence, QFT, brain) one *must* use all symmetries of the problem. Chaotic dynamics is no easier to grasp than quantum mechanics. ChaosBook approaches both with the same set of tools, some of them group-theoretical.

[appendix A1.6](#)



The erudite reader might prefer to skip the lengthy group-theoretic overture and go directly to D_1 example [11.3](#), Lorenz flow example [11.8](#), and D_3 example [11.5](#), backtrack as needed.

10.1 Discrete symmetries



We show that a symmetry equates multiplets of equivalent orbits, or ‘stratifies’ the state space into equivalence classes. We start by defining a finite (discrete) group, its state space representations, and what we mean by a *symmetry* (*invariance* or *equivariance*) of a dynamical system. As is always the problem with ‘gruppenpest’ (read appendix A1.6) way too many abstract notions have to be defined before an intelligent conversation can take place. Perhaps best to skim through this section on the first reading, then return to it later as needed.



Definition: A **group** consists of a set of elements

$$G = \{e, g_2, \dots, g_n, \dots\} \quad (10.1)$$

and a group multiplication rule $g_j \circ g_i$ (often abbreviated as $g_j g_i$), satisfying



1. Closure: If $g_i, g_j \in G$, then $g_j \circ g_i \in G$
2. Associativity: $g_k \circ (g_j \circ g_i) = (g_k \circ g_j) \circ g_i$
3. Identity e : $g \circ e = e \circ g = g$ for all $g \in G$
4. Inverse g^{-1} : For every $g \in G$, there exists a unique element $h = g^{-1} \in G$ such that $h \circ g = g \circ h = e$.

If the group is finite, the number of elements, $|G| = n$, is called the *order* of the group.

The theory of finite groups is developed on two levels. There is a beautiful theory of *groups* as abstract entities which yields the classification of their structures and their irreducible, orthogonal representations in terms of characters. Then there is the considerably messier matter of *group representations*. For example, we are concerned with ways that a given symmetry group acts on and stratifies a particular state space during symmetry reduction and block-diagonalizable quantum mechanics. What helps us here is that symmetries ‘commute’ with dynamics, i.e., we can first reduce a given state space to its irreducible components, using symmetry alone, and then study the action of dynamics on these subspaces. As our intuition is based on physical manifestations of group actions, in this brief review we shall freely switch gears between the abstract and the representation levels whenever pedagogically convenient.

Whatever else you must do, do work through example 11.5. Once you understand how this works out for the symmetries of an equilateral triangle, or, equivalently, for the three disk billiard of figure 10.1, you know almost everything you need to know about general, non-abelian finite groups.



example 10.1
p. 188



example 10.2
p. 188

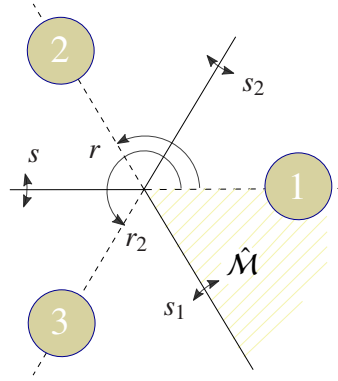


Figure 10.1: Dihedral group D_3 of symmetries of three disks on an equilateral triangle. A fundamental domain $\hat{M}$ indicated by the shaded wedge. Work through example 11.5.

Definition: Coordinate transformations. Consider a map $x' = f(x)$, $x, x' \in \mathcal{M}$. An *active* coordinate transformation Mx corresponds to a non-singular $[d \times d]$ matrix M that maps the initial vector $x \in \mathcal{M}$ onto another vector $Mx \in \mathcal{M}$. The corresponding *passive* coordinate transformation $x' \rightarrow M^{-1}x'$ changes the coordinate system with respect to which the final vector $x' \in \mathcal{M}$ is measured. Together,



$$\hat{f}(x) = M^{-1}f(Mx). \quad (10.2)$$

(For general nonlinear coordinate transformations, see Appendix A2.)

Definition: Matrix group. The set of $[d \times d]$ -dimensional real non-singular matrices $A, B, C, \dots \in GL(d)$ acting in a d -dimensional vector space $V \in \mathbb{R}^d$ forms the general linear group $GL(d)$ under matrix multiplication. The product of matrices A and B gives the matrix C , $Cx = B(Ax) = (BA)x \in V$, for all $x \in V$. The unit matrix $\mathbf{1}$ is the identity element which leaves all vectors in V unchanged. Every matrix in the group has a unique inverse.

Definition: Matrix representation. Linear action of a group element g on states $x \in \mathcal{M}$ is given by a finite non-singular $[d \times d]$ matrix $D(g)$, the *matrix representation* of element $g \in G$. For brevity we shall often denote by ' g ' both the abstract group element and its matrix representation, $D(g)x \rightarrow gx$.

However, when dealing simultaneously with several representations of the same group action, the notation $D^{(\mu)}(g)$ is preferable, where μ is a representation label (see appendix A10.1). A linear or matrix representation $D(G)$ of the abstract group G acting on a *representation space* V is a group of matrices $D(G)$ such that

1. Any $g \in G$ is mapped to a matrix $D(g) \in D(G)$.
2. The group product $g_2 \circ g_1$ is mapped onto the matrix product $D(g_2 \circ g_1) = D(g_2)D(g_1)$.

3. The associativity follows from the associativity of matrix multiplication, $D(g_3 \circ (g_2 \circ g_1)) = D(g_3)(D(g_2)D(g_1)) = (D(g_3)(D(g_2)))D(g_1)$.
4. The identity element $e \in G$ is mapped onto the unit matrix $D(e) = 1$ and the inverse element $g^{-1} \in G$ is mapped onto the inverse matrix $D(g^{-1}) = D(g)^{-1}$.

Several 3D representations of the group of order 2 are given in example 10.4.



example 10.3
p. 188



example 10.4
p. 188

If the coordinate transformation g belongs to a linear non-singular representation of a discrete finite group G , for any element $g \in G$ there exists a number $m \leq |G|$ such that

$$g^m \equiv \underbrace{g \circ g \circ \cdots \circ g}_{m \text{ times}} = e \quad \rightarrow \quad |\det D(g)| = 1. \quad (10.3)$$

As the modulus of its determinant is unity, $\det g$ is an m th root of 1. This is the reason why all finite groups have unitary representations.

Definition: Symmetry of a dynamical system. This is so important that we will define it three times:

1. A group G is a *symmetry* of the dynamics if for every solution $f(x) \in \mathcal{M}$ and $g \in G$, $gf(x)$ is also a solution. 
2. Another way to state this: A dynamical system $(\mathcal{M}, f)$ is *invariant* (or *G-equivariant*) under a symmetry group G if the time evolution $f : \mathcal{M} \rightarrow \mathcal{M}$ (a discrete time map f , or the continuous flow f^t map from the d -dimensional manifold $\mathcal{M}$ into itself) commutes with all actions of G ,

$$f(gx) = gf(x). \quad (10.4)$$

3. In the language of physicists: The ‘law of motion’ is invariant, i.e., retains its form in any symmetry-group related coordinate frame (10.2),

$$f(x) = g^{-1}f(gx), \quad (10.5)$$

for $x \in \mathcal{M}$ and any finite non-singular $[d \times d]$ matrix representation g of element $g \in G$. As this is true for any state x , one can state this more compactly as $f \circ g = g \circ f$, or $f = g^{-1} \circ f \circ g$.

Why ‘equivariant?’ A scalar function $h(x)$ is said to be *G-invariant* if $h(x) = h(gx)$ for all $g \in G$. The group actions map the solution $f : \mathcal{M} \rightarrow \mathcal{M}$ into different (but equivalent) solutions $gf(x)$, hence the invariance condition $f(x) = g^{-1}f(gx)$ appropriate to vectors (and, more generally, tensors). The full set of such solutions is *G-invariant*, but the flow that generates them is said to be *G-equivariant*.

It is obvious from the context, but for verbal emphasis applied mathematicians like to distinguish the two cases by *in/equi*-variant. The distinction is helpful in distinguishing the dynamics written in the original, equivariant coordinates from the dynamics rewritten in terms of *invariant* coordinates, see sects. 11.5 and 13.2.



example 10.5
p. 188



example 10.6
p. 189



example 10.10
p. 190

10.2 Subgroups, cosets, classes

Normal is just a setting on a washing machine.
—Borgette, Borgo's daughter

Inspection of figure 11.1 indicates that various 3-disk orbits are the same up to a symmetry transformation. Here we set up some group-theoretic notions needed to describe such relations. The reader might prefer to skip to sect. 11.1, backtrack as needed.



Definition: Subgroup. A set of group elements $G_p = \{e, b_2, b_3, \dots, b_h\} \subseteq G$ closed under group multiplication forms a subgroup of order $h = |G_p|$.

Definition: Coset. The set of h elements $\{c, cb_2, cb_3, \dots, cb_h\}$, $c \in G$ but not in G_p , is called *left coset* cG_p . A subgroup G_p thus partitions the G group elements into G_p its $m_p - 1$ cosets.

Definition: Index. For a finite discrete group, the number of left cosets

$$m_p = |G|/|G_p| \quad (10.6)$$

is called the *index* of subgroup G_p . For reasons explained in remark 11.3, here we shall refer to index m_p as *multiplicity*.

Lagrange's Theorem. A coset cannot form a subgroup, since it does not include the identity element. A nontrivial subgroup G_p can exist only if $|G|$, the order of the group, is divisible by $|G_p|$.



example 10.8
p. 189

'Class' is a notion that says the three 3-disk 2-cycles in figure 11.1 are one and the same solution, up to a coordinate change.

Definition: Class. An element $b \in G$ is *conjugate* to a if $b = c a c^{-1}$, where c is some other group element. If b and c are both conjugate to a , they are conjugate to each other. Application of all conjugations separates the set of group elements into mutually non-conjugate subsets called *classes*, *types* or *conjugacy classes*. The identity e is always in the class $\{e\}$ of its own. This is the only class which is a subgroup, and all other classes lack the identity element.

exercise 10.1

exercise 10.5



example 10.9
p. 190

The geometrical significance of classes is clear from (10.5); it is the way co-ordinate transformations act on mappings. The action, such as a reflection or rotation, of an element is equivalent to redefining the coordinate frame.

Definition: Conjugate symmetry subgroups. The splitting of a group G into a symmetry group G_p of orbit $\mathcal{M}_p$ and $m_p - 1$ cosets cG_p relates the orbit $\mathcal{M}_p$ to $m_p - 1$ other distinct orbits $c\mathcal{M}_p$. All of them have equivalent symmetry subgroups, or, more precisely, the points on the same group orbit have *conjugate symmetry subgroups* (or *conjugate stabilizers*):

remark 11.3
exercise 10.2

$$G_{c p} = c G_p c^{-1}, \quad (10.7)$$

i.e., if G_p is the symmetry of orbit $\mathcal{M}_p$, elements of the coset space $c \in G/G_p$ generate the $m_p - 1$ distinct copies of $\mathcal{M}_p$.

Definition: Invariant subgroup. A subgroup $G_p \subseteq G$ is an *invariant* subgroup or *normal divisor* if it consists of complete classes. Class is complete if no conjugation takes an element of the class out of G_p .

Think of action of G_p within each coset as identifying its $|G_p|$ elements as equivalent. This leads to the notion of the *factor group* or *quotient group* G/G_p of G , with respect to the invariant subgroup G_p . G_p thus divides G into G_p and $m - 1$ cosets, each of order $|G_p|$. The order of G/G_p is $m = |G|/|G_p|$, and its multiplication table can be worked out from the G multiplication table class by class, with the subgroup G_p playing the role of identity. G/G_p is *homeomorphic* to G , with $|G_p|$ elements in a class of G represented by a single element in G/G_p .

10.3 Orbits, quotient space

section 2.1

Definition: Orbit. The subset $\mathcal{M}_{x_0} \subset \mathcal{M}$ traversed by the infinite-time trajectory of a given point x_0 is called the *orbit* (or *time orbit*, or *solution*) $x(t) = f^t(x_0)$. An orbit is a *time-invariant* notion: it refers to the set of all states that can be reached in time from x_0 . Thus, as a set it, is invariant under time evolution. The full state space $\mathcal{M}$ is a union of such orbits. We label a generic orbit $\mathcal{M}_{x_0}$ by any point belonging to it, $x_0 = x(0)$ for example.

A generic orbit might be ergodic, unstable and essentially uncontrollable. The ChaosBook strategy is to populate the state space by a hierarchy of orbits which are *compact invariant sets* (equilibria, periodic orbits, invariant tori, ...), each computable in a finite time. They are a set of zero Lebesgue measure, but dense on the non-wandering set, and are to a generic orbit what fractions are to normal numbers on the unit interval. We label orbits confined to compact invariant sets by whatever alphabet we find convenient in a given context: point $EQ = x_{EQ} = \mathcal{M}_{EQ}$ for an equilibrium, 1-dimensional loop $p = \mathcal{M}_p$ for a prime periodic orbit p , etc. (note also discussion on page 224, and the distinction between trajectory and orbit made in sect. 2.1; a trajectory is a finite-time segment of an orbit).

Definition: Group orbit or the G -orbit of the point $x \in \mathcal{M}$ is the set

$$\mathcal{M}_x = \{g x \mid g \in G\} \quad (10.8)$$

of all state space points into which x is mapped under the action of G . If G is a symmetry, intrinsic properties of an equilibrium (such as stability eigenvalues) or a cycle p (period, Floquet multipliers) evaluated anywhere along its G -orbit are the same.

section 13.2

A symmetry thus reduces the number of inequivalent solutions $\mathcal{M}_p$. So we also need to describe the symmetry of a *solution*, as opposed to (10.5), the symmetry of the *system*.

Definition: Reduced state space. The action of group G partitions the state space $\mathcal{M}$ into a union of group orbits. This set of group orbits, denoted $\mathcal{M}/G$, has many names: *reduced state space*, *quotient space* or any of the names listed on page 237.

Definition: Fundamental domain. The images of a single point x under all actions of a discrete group G form a G -orbit $\mathcal{M}_x$. A fundamental domain $\hat{\mathcal{M}} = \mathcal{M}/G$ is a subset of the state space $\mathcal{M}$, which contains exactly one point from each G -orbit. It is an explicit state space realization of the abstract notion of the reduced state space $\mathcal{M}/G$, in the case that G is a discrete group.

A fundamental domain can be defined in different ways, here exemplified by figures 10.1, 11.1, 11.5, 11.3, 11.2 (b) and 24.3. Ideally it is a connected subset with restrictions on its boundary and ensures that no points are double-counted. The set of images of a fundamental domain under the group action then tiles the entire state space.

Reduction of the dynamical state space is discussed in sect. 11.3 for discrete symmetries, and in sect. 13.2 for continuous symmetries.

Definition: Fixed-point subspace. $\mathcal{M}_{G_p}$ is the set of all state space points left G_p -fixed, point-wise invariant under subgroup or ‘centralizer’ $G_p \subseteq G$ action

$$\mathcal{M}_{G_p} = \text{Fix}(G_p) = \{x \in \mathcal{M} \mid hx = x \text{ for all } h \in G_p\}. \quad (10.9)$$

Points in state space subspace $\mathcal{M}_G$, which are fixed points of the full group action, are called *invariant points*,

$$\mathcal{M}_G = \text{Fix}(G) = \{x \in \mathcal{M} \mid gx = x \text{ for all } g \in G\}. \quad (10.10)$$

Definition: Flow invariant subspace. A typical point in fixed-point subspace $\mathcal{M}_{G_p}$ moves with time, but, due to equivariance (10.4), its trajectory $x(t) = f^t(x)$ remains within $f(\mathcal{M}_{G_p}) \subseteq \mathcal{M}_{G_p}$ for all times,

$$hf^t(x) = f^t(hx) = f^t(x), \quad h \in G_p. \quad (10.11)$$

Hence, what happens in *flow invariant subspace*, stays in flow invariant subspace. This suggests a systematic approach to seeking compact invariant solutions. The larger the symmetry subgroup, the smaller $\mathcal{M}_{G_p}$, easing the numerical searches. For this reason, we recommend starting with the largest subgroups G_p first.

We can often decompose the state space into smaller subspaces, with group acting within each ‘chunk’ separately:

Definition: Invariant subspace. $\mathcal{M}_\alpha \subset \mathcal{M}$ is an *invariant* subspace if

$$\{\mathcal{M}_\alpha \mid gx \in \mathcal{M}_\alpha \text{ for all } g \in G \text{ and } x \in \mathcal{M}_\alpha\}. \quad (10.12)$$

$\{0\}$ and $\mathcal{M}$ are always invariant subspaces. So is any $\text{Fix}(G_p)$ which is point-wise invariant under action of G .

Definition: Irreducible subspace. A space $\mathcal{M}_\alpha$, whose only invariant subspaces under the action of G are $\{0\}$ and $\mathcal{M}_\alpha$, is called *irreducible*.

Definition: Reducibility. If state space $\mathcal{M}$ on which G acts can be written as a direct sum of irreducible subspaces, then the representation of G on state space $\mathcal{M}$ is completely reducible.

This being group theory, definitions could go on forever. But we stop here, hopefully having defined everything that we need at the moment, and we pile on a few more definitions in sect. 11.1, chapter 12, chapter 25, chapter 26, chapter 30, and appendix A10. If you desire to go beyond that, we recommend the $n \rightarrow \infty$ group theory textbooks.

Résumé

A group G is a *symmetry* of the dynamical system (M, f) , if its ‘law of motion’ retains its form under all symmetry-group actions, $f(x) = g^{-1}f(gx)$. A mapping f is said to be *invariant* if $gf = f$, where g is any element of G . If the mapping and the group actions commute, $gf = fg$, f is said to be *equivariant*. The governing dynamical equations are equivariant with respect to the symmetry group G .

Commentary

Remark 10.1. Literature. We found Tinkham [16] the most enjoyable no-nonsense, user-friendly introduction to basic concepts. Slightly longer, but perhaps student-friendlier is *Part I Basic Mathematics* by Dresselhaus *et al.* [5]. Byron and Fuller [1], the last chapter of volume two, offers an introduction even more compact than Tinkham’s. For a summary of the theory of discrete groups see, for example, Johnson [10]. Chapter 3 of Rebecca Hoyle [9] is the most student-friendly overview of group theory that a nonlinear dynamicist might need. However, quotient groups and the reduction of dynamics to a fundamental domain is not discussed at all in this chapter. For those topics, [Fundamental domain](#) wiki and [Quotient group](#) wiki are clear and very helpful. Curiously, we have not read any of the group theory books that Hoyle recommends as background reading, which just confirms that there are way too many group theory books out there. For example, one that you will not find useful at all is ref. [3]. The reason is presumably that 20th century physics (which motivated much of the work on the modern group theory) focused on linear representations used in quantum mechanics, crystallography and quantum field theory. We shall need these techniques in Chapter 25, where we reduce the linear action of evolution operators to irreducible subspaces. However, in ChaosBook we are looking at nonlinear dynamics, and the emphasis is on the symmetries of orbits, their reduced state space sisters, and the isotypic decomposition of their linear stability matrices.

appendix A1.6

In ChaosBook, we focus on chaotic dynamics and skirt around the theory of bifurcations, which is the landscape between the boredom of regular motion and the thrills of chaos. Landau [11] was the first to discuss the role symmetries play in constraining types of possible bifurcations, in the context of weak nonlinear theory and fluid flow instabilities. Chapter 4 of Rebecca Hoyle [9] is a student-friendly introduction to the treatment of bifurcations in the presence of symmetries, worked out in full detail and generality in monographs by Golubitsky, Stewart and Schaeffer [7], Golubitsky and Stewart [6] and Chossat and Lauterbach [2]. Sartori [14] Sect. 1.3 offers a concise summary of group-theoretical definitions. Chap. 8 of Govaerts [8] reviews numerical methods that employ equivariance with respect to compact, and mostly discrete groups. (continued in remark 12.1)

References

- [1] F. W. Byron and R. W. Fuller, [Mathematics of Classical and Quantum Physics](#) (Dover, New York, 1992).

- [2] P. Chossat and R. Lauterbach, *Methods in Equivariant Bifurcations and Dynamical Systems* (World Scientific, Singapore, 2000).
- [3] P. Cvitanović, *Group Theory: Birdtracks, Lie's and Exceptional Groups* (Princeton Univ. Press, Princeton NJ, 2008).
- [4] C. Dong and H. Liu, “Unstable cycles for the Burke-Shaw system via variational approach”, *Int. J. Mod. Phys. B* **33**, 1950240 (2019).
- [5] M. S. Dresselhaus, G. Dresselhaus, and A. Jorio, *Group Theory: Application to the Physics of Condensed Matter* (Springer, New York, 2007).
- [6] M. Golubitsky and I. Stewart, *The Symmetry Perspective* (Birkhäuser, Boston, 2002).
- [7] M. Golubitsky, I. Stewart, and D. G. Schaeffer, *Singularities and Groups in Bifurcation Theory*, Vol. 2 (Springer, New York, 1988).
- [8] W. J. F. Govaerts, *Numerical Methods for Bifurcations of Dynamical Equilibria* (SIAM, Philadelphia, 2000).
- [9] R. Hoyle, *Pattern Formation: An Introduction to Methods* (Cambridge Univ. Press, Cambridge UK, 2006).
- [10] S. G. Johnson, *Topics in applied mathematics*, MIT course 18.325, 2005.
- [11] L. D. Landau, “On the problem of turbulence”, *Dokl. Akad. Nauk SSSR* **44**, 339 (1944).
- [12] C. Letellier, P. Dutertre, J. Reizner, and G. Gouesbet, “Evolution of a multimodal map induced by an equivariant vector field”, *J. Phys. A* **29**, 5359–5373 (1996).
- [13] C. Letellier, G. B. T. Tsankov, and R. Gilmore, “Large-scale structural reorganization of strange attractor”, *Phys. Rev. E* **72**, 026212 (2005).
- [14] G. Sartori, “Geometric invariant theory: a model-independent approach to spontaneous symmetry and/or supersymmetry breaking”, *Rivista Nuovo Cim.* **14**, 1–120 (1991).
- [15] R. Shaw, “Strange attractors, chaotic behavior, and information flow”, *Z. Naturf. A* **36**, 80–112 (1981).
- [16] M. Tinkham, *Group Theory and Quantum Mechanics* (Dover, New York, 2003).

10.4 Examples

Example 10.1. Finite groups. Some finite groups that frequently arise in applications:

- C_n : the *cyclic group* of order n .
- D_n : the *dihedral group* of order $2n$, rotations and reflections in plane that preserve a regular n -gon.
- S_n : the *symmetric group* of all permutations of n symbols, order $n!$.

[return to p. 179](#)

Example 10.2. Cyclic and dihedral groups. The cyclic group $C_n \subset \text{SO}(2)$ of order n is generated by one element. For example, this element can be rotation through $2\pi/n$.

The dihedral group $D_n \subset \text{O}(2)$, $n > 2$, can be generated by two elements one at least of which must reverse orientation. For example, take σ corresponding to reflection in the x -axis. $\sigma^2 = e$; such operation σ is called an *involution*. r to rotation through $2\pi/n$, then $D_n = \langle \sigma, C \rangle$, and the defining relations are $\sigma^2 = r_n = e$, $(r\sigma)^2 = e$.

[return to p. 179](#)

Example 10.3. Discrete groups of order 2 on $\mathbb{R}^3$. Three types of discrete group of order 2 can arise by linear action on our 3-dimensional Euclidean space $\mathbb{R}^3$:



$$\begin{aligned} \text{reflections: } s(x, y, z) &= (x, y, -z) \\ \text{rotations: } r(x, y, z) &= (-x, -y, z) \\ \text{inversions: } P(x, y, z) &= (-x, -y, -z). \end{aligned} \quad (10.13)$$

s is a *reflection* (or an inversion) through the $[x, y]$ plane. r is $[x, y]$ -plane, constant z *rotation* by π about the z -axis (or an inversion thorough the z -axis). P is an *inversion* (or parity operation) through the point $(0, 0, 0)$. Singly, each operation generates a group of order 2: $D_1 = \{e, s\}$, $C_2 = \{e, r\}$, and $D_1 = \{e, P\}$. Together, they form the dihedral group $D_2 = \{e, s, r, P\}$ of order 4. (continued in example 10.4)

[return to p. 181](#)

Example 10.4. Discrete operations on $\mathbb{R}^3$. (Continued from example 10.3) The matrix representation of reflections, rotations and inversions defined by (10.13) is

$$D(s) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad D(r) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad D(P) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad (10.14)$$

with $\det D(r) = 1$, $\det D(s) = \det D(P) = -1$; that is why we refer to r as a rotation, and s, P as inversions. As $g^2 = e$ in all three cases, these are groups of order 2. (continued in example 10.6)

[return to p. 181](#)

Example 10.5. A reflection symmetric 1d map. Consider a 1d map f with reflection symmetry $f(-x) = -f(x)$, such as the bimodal ‘sawtooth’ map of figure 10.2, piecewise-linear on the state space $\mathcal{M} = [-1, 1]$, a compact 1-dimensional line interval, split into three regions $\mathcal{M} = \mathcal{M}_L \cup \mathcal{M}_C \cup \mathcal{M}_R$. Denote the reflection operation by $sx = -x$. The



Figure 10.2: The bimodal Ulam sawtooth map with the D_1 symmetry $f(-x) = -f(x)$. If the trajectory (a) $x_0 \rightarrow x_1 \rightarrow x_2 \rightarrow \dots$ is a solution, so is its reflection (b) $sx_0 \rightarrow sx_1 \rightarrow sx_2 \rightarrow \dots$. (work through example 10.5; continued in figure 11.4).

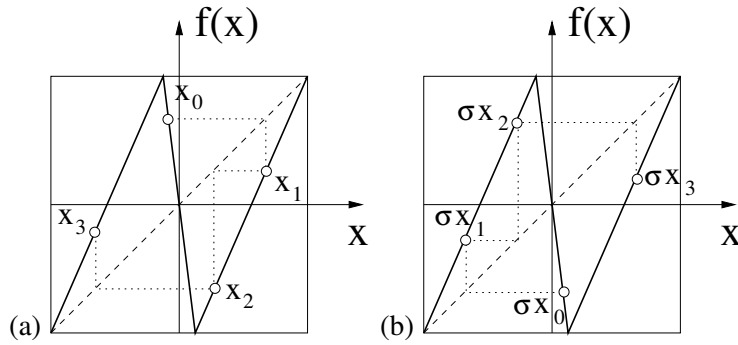


Table 10.1: D_3 group and class operator multiplication tables.

D_3	1	r	r_2	s	s_1	s_2
1	1	r	r_2	s	s_1	s_2
r	r	r_2	1	s_2	s	s_1
r_2	r_2	1	r	s_1	s_2	s
s	s	s_1	s_2	1	r	r_2
s_1	s_1	s_2	s	r_2	1	r
s_2	s_2	s	s_1	r	r_2	1

2-element group $G = \{e, s\}$ goes by many names, such as Z_2 or C_2 . Here we shall refer to it as D_1 , dihedral group generated by a single reflection. The G -equivariance of the map implies that if $\{x_n\}$ is a trajectory, than also $\{sx_n\}$ is a symmetry-equivalent trajectory because $sx_{n+1} = sf(x_n) = f(sx_n)$ (continued in example 11.3)

[return to p. 182](#)

Example 10.6. Equivariance of the Lorenz flow. (Continued from example 10.4) The velocity field in Lorenz equations (2.25)

[exercise 10.3](#)

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} \sigma(y - x) \\ \rho x - y - xz \\ xy - bz \end{bmatrix} \quad (10.15)$$

is equivariant under the action of cyclic group $C_2 = \{e, r\}$ acting on $\mathbb{R}^3$ by a π rotation about the z axis,

$$r(x, y, z) = (-x, -y, z). \quad (10.16)$$

(continued in example 11.8)

[return to p. 182](#)

Example 10.7. The group multiplication table for D_3 : See table 10.1.

Example 10.8. Subgroups, cosets of D_3 . (Continued from example 11.6)

The 3-disks symmetry group, the D_3 dihedral group (11.8) has six subgroups

$$\{e\}, \quad \{e, s\}, \{e, s_1\}, \{e, s_2\}, \quad \{e, r, r_2\}, \quad D_3. \quad (10.17)$$

The left cosets of subgroup $D_1 = \{e, s\}$ are $\{r, s_1\}, \{r_2, s_2\}$. The coset of subgroup $C_3 = \{e, r, r_2\}$ is $\{s, s_1, s_2\}$. The significance of the coset is that if a solution has a symmetry H , for example the symmetry of a 3-cycle $\overline{123}$ is C_3 , then all elements in a coset act on it the same way, for example $\{s, s_1, s_2\}\overline{123} = \overline{132}$.

The nontrivial subgroups of D_3 are $D_1 = \{e, \sigma\}$, consisting of the identity and any one of the reflections, of order 2, and $C_3 = \{e, r, r_2\}$, of order 3, so possible cycle multiplicities are $|G|/|G_p| = 1, 2, 3$ or 6. Only the fixed point at the origin has full symmetry $G_p = G$. Such equilibria exist for smooth potentials, but not for the 3-disk billiard. Examples of other multiplicities are given in figure 11.1 and figure 11.6. (continued in example 10.9)

[return to p. 182](#)

Example 10.9. Classes of D_3 . (Continued from example 10.8)

The three classes of the 3-disk symmetry group $D_3 = \{e, r, r_2, s, s_1, s_2\}$, are the identity, any one of the reflections, and the two rotations,

$$\{e\}, \quad \left\{ \begin{matrix} s \\ s_1 \\ s_2 \end{matrix} \right\}, \quad \left\{ \begin{matrix} r \\ r_2 \end{matrix} \right\}. \quad (10.18)$$

In other words, the group actions either flip or rotate. (continued in example 11.7)

[return to p. 183](#)

Example 10.10. Discrete symmetries of the plane Couette flow. The plane Couette flow is a fluid flow bounded by two countermoving planes, in a cell periodic in streamwise and spanwise directions. The Navier-Stokes equations for the plane Couette flow have two discrete symmetries: reflection through the (streamwise , wall-normal) plane, and rotation by π in the (streamwise , wall-normal) plane. That is why the system has equilibrium and periodic orbit solutions, as well as relative equilibrium and relative periodic orbit solutions discussed in chapter 12). They belong to discrete symmetry subspaces. (continued in example 12.2)

[return to p. 182](#)

10.5 Exercises

- 10.1. **Transitivity of conjugation.** Assume that $g_1, g_2, g_3 \in G$ and both g_1 and g_2 are conjugate to g_3 . Prove that g_1 is conjugate to g_2 .
- 10.2. **Isotropy subgroup of gx .** Prove that for $g \in G$, x and gx have conjugate isotropy subgroups:

$$G_{gx} = g G_x g^{-1}$$

- 10.3. **C_2 -equivariance of Lorenz system.** Verify that the vector field in Lorenz equations (10.15)

$$\dot{x} = v(x) = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} \sigma(y - x) \\ \rho x - y - xz \\ xy - bz \end{bmatrix} \quad (10.19)$$

is equivariant under the action of cyclic group $Z_2 = \{e, r\}$ acting on $\mathbb{R}^3$ by a π rotation about the z axis,

$$r(x, y, z) = (-x, -y, z),$$

as claimed in example 10.6. (continued in exercise 11.4)

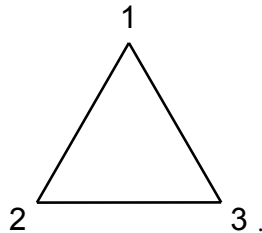
- 10.4. **C_2 -equivariance of Burke-Shaw system.** The Burke-Shaw system [4, 12, 13, 15] is a close relative of the Lorenz system:

$$\dot{x} = v(x) = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} -s(x + y) \\ -y - sxxz \\ sxy + v \end{bmatrix} \quad (10.20)$$

- (a) Plot a long-time simulation [4, 12] for $(s, v) = (10.5, 4.272)$.
- (b) Verify that the Burke-Shaw equations are equivariant under the action of the cyclic group $C_2 = \{e, r\}$ acting on $\mathbb{R}^3$ by a π rotation about the z axis,

$$r(x, y, z) = (-x, -y, z).$$

- 10.5. **D_3 : symmetries of an equilateral triangle.** Consider the group D_3 , the symmetry group of an equilateral triangle:



- (a) List the group elements and the corresponding geometric operations
- (b) Find the subgroups of the group D_3 .
- (c) Find the classes of D_3 and the number of elements in them, guided by the geometric interpretation of group elements. Verify your answer using the definition of a class.
- (d) List the conjugacy classes of subgroups of D_3 . (continued as exercise 12.2 and exercise 25.3)

Chapter 11

World in a mirror

Even the butterfly that started the hurricane flapped its wings for a reason.

— Louis Menand, *Thinking Sideways*, New Yorker, 30 March 2015

SO FAR WE HAVE discussed the structure of a group as an abstract entity. Now we switch gears and describe the action of the group on state space. This is the key step; if a set of solutions is equivalent by symmetry (let's say they live on a circle), we would like to represent it by a single solution (cut the circle at a point, or rewrite the dynamics in a 'symmetry reduced state space', where the circle of equivalent solutions is represented by a single state space point). In this chapter we study quotienting of discrete symmetries, and in chapter 12 we study symmetry reduction for continuous symmetries. We look at individual orbits, and the ways they are interrelated by symmetries. This sets the stage for a discussion of how symmetries affect global densities of trajectories, and the factorization of spectral determinants to be undertaken in chapter 25.

As we shall show here and in chapter 25, discrete symmetries simplify the dynamics in quite a beautiful way. If dynamics is invariant under a set of discrete symmetries G , the state space $\mathcal{M}$ is tiled by a set of symmetry-related tiles, and the dynamics can be reduced to dynamics within one such tile, the *fundamental domain* $\mathcal{M}/G$. In the presence of symmetry, the notion of a prime periodic orbit has to be reexamined. Within this context, a set of symmetry-related full state space cycles is often replaced by a much shorter *relative periodic orbit*, the shortest segment of the full state space cycle which tiles the cycle and all of its copies under the action of the group. Furthermore, the group operations that relate distinct tiles do double duty as letters of an alphabet which assigns symbolic itineraries to trajectories.

section 14.1

11.1 Symmetries of solutions

Solutions of an equivariant system can satisfy all of a system's symmetries, a subgroup of them, or have no symmetry at all. For a generic ergodic orbit $f^t(x)$, the trajectory and any of its images under action of $g \in G$ are distinct with probability one, $f^t(x) \cap g f^{t'}(x) = \emptyset$ for all t, t' . For example, a typical turbulent trajectory of pipe flow has no symmetry beyond the identity, so its symmetry group is the trivial subgroup $\{e\}$. For compact invariant sets, such as fixed points and periodic orbits, the situation is very different. For example, the symmetry of the laminar solution of plane Couette flow is the full symmetry of its Navier-Stokes equations. In between we find solutions whose symmetries are subgroups of the full symmetry of dynamics.

Definition: Isotropy subgroup. The maximal set of group actions which maps a state space point x into itself,

$$G_x = \{g \in G : gx = x\}, \quad (11.1)$$

is called the *isotropy group* (or *stability subgroup* or *little group*) of x . Think of a point $(0, 0, z)$, $z \neq 0$ on z axis in 3 dimensions. Its isotropy group is the $O(2)$ group of rotations in the $\{x, y\}$ plane.

A solution usually exhibits less symmetry than the equations of motion. The symmetry of a solution is thus a subgroup of the symmetry group of dynamics. We thus also need a notion of *set-wise* invariance, as opposed to the above *point-wise* invariance under G_x .

exercise 11.1

Definition: Symmetry of a solution. We shall refer to the maximal subgroup $G_p \subseteq G$ of actions on state space points within a compact set $\mathcal{M}_p$. These actions leave no point fixed but leave the set invariant, as the *symmetry* G_p of the solution labelled p ,

$$G_p = \{g \in G_p \mid gx \in \mathcal{M}_p, gx \neq x \text{ for } g \neq e\}, \quad (11.2)$$

and reserve the notion of 'isotropy' of a set $\mathcal{M}_p$ for the subgroup G_p that leaves each point in it fixed.

A cycle p is G_p -*symmetric* (*set-wise symmetric*, *self-dual*), if the action of elements of G_p on the set of periodic points $\mathcal{M}_p$ reproduces the set. $g \in G_p$ acts as a shift in time, mapping the periodic point $x \in \mathcal{M}_p$ into another periodic point.



example 11.1
p. 204

Definition: Multiplicity. For a finite group, the index (10.6) $m_p = |G|/|G_p|$ of orbit p 's symmetry group G_p is its multiplicity, the number of p 's equivalent copies.

remark 11.3

Definition: Stratum. A stratum is the union of group orbits of the same type. Two orbits p, p' belong to the same stratum if and only if their symmetries $G_p, G_{p'}$ are conjugate. In other words, a stratum is to state space what a class is to the set of all group elements in G .

Definition: G_p -fixed orbits: An equilibrium x_q or a compact solution p is point-wise or G_p -fixed, if it lies in the invariant points subspace $\text{Fix}(G_p)$, $gx = x$ for all $g \in G_p$, and $x = x_q$ or $x \in \mathcal{M}_p$. A solution that is G -invariant under all group G operations has multiplicity 1. Stability of such solutions will have to be examined with care, as they lie on the boundaries of domains related by the action of the symmetry group.

The symmetry group of a solution is often called a *stabilizer* or an *isotropy subgroup*. Saying that G_p is the symmetry of the solution p , or that the orbit $\mathcal{M}_p$ is ' G_p -invariant', accomplishes as much without confusing you with all these names (see remark 10.1). In what follows we say "the symmetry of the periodic orbit p is $Z_2 = \{e, R\}$," rather than bandy about 'stabilizers' and such.

The key concept in the classification of dynamical orbits is their symmetry. We note three types of solutions: (i) fully asymmetric solutions a , (ii) subgroup $G_{\tilde{s}}$ set-wise invariant cycles s built by repeats of relative cycle segments $\tilde{s}$, and (iii) isotropy subgroup G_{EQ} -invariant equilibria or point-wise G_p -fixed cycles b . These are illustrated in figures of examples 11.1 to 11.3.

 example 11.2
p. 204

Definition: Asymmetric (or fully asymmetric) orbits. An orbit (in particular, an equilibrium or periodic orbit) has no symmetry if $\{x_a\} \cap \{gx_a\} = \emptyset$ for any $g \in G$, where $\{x_a\}$ is the set of periodic points belonging to cycle a . Thus $g \in G$ generate $|G|$ distinct orbits with the same number of points and the same stability properties.

 example 11.3
p. 204

 example 11.5
p. 205

 example 11.6
p. 205

In example 11.7, we illustrate the non-abelian, noncommutative group structure of the 3-disk game of pinball of sect. 1.3, which has symmetry group elements that do not commute.

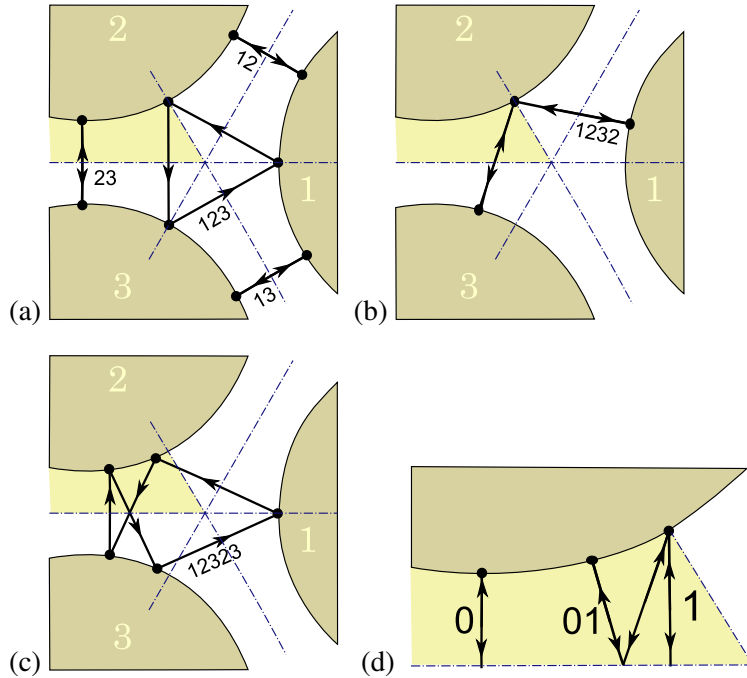
exercise 10.5

 example 11.7
p. 206

exercise 11.4
exercise 11.5

Consider next perhaps the simplest 3-dimensional flow with a symmetry, the iconic flow of Lorenz of figure 11.2(a). The example is long but worth working through: the symmetry-reduced dynamics is much simpler than the original Lorenz flow.

Figure 11.1: The 3-disk pinball cycles: (a) $\overline{12}$, $\overline{13}$, $\overline{23}$, $\overline{123}$; the clockwise $\overline{132}$ not drawn. (b) Cycle $\overline{1232}$; the symmetry related $\overline{1213}$ and $\overline{1323}$ not drawn. (c) Cycle $\overline{12323}$; cycles $\overline{12123}$, $\overline{12132}$, $\overline{12313}$, $\overline{13131}$ and $\overline{13232}$ not drawn. (d) The fundamental domain, i.e., the light-shaded $1/6$ th wedge in (a), consisting of a section of a disk, two segments of symmetry axes acting as straight mirror walls, and the escape gap to the left. The above 14 full-space cycles restricted to the fundamental domain and recoded in binary reduce to the two fixed points $\overline{0}$, $\overline{1}$, 2-cycle $\overline{10}$, and 5-cycle $\overline{00111}$ (not drawn). See figure 11.3 for the $\overline{001}$ cycle. Work through example 11.6.



example 11.8
p. 206



example 11.9
p. 207

Note: nonlinear coordinate transformations such as the doubled-polar angle representation (11.13) and figure 11.2 (b) are *not* required to implement the symmetry quotienting M/G . We deploy them only as a visualization aid that might help the reader disentangle 2-dimensional projections of higher-dimensional flows. All numerical calculations can still be carried in the initial, full state space formulation of a flow, with symmetry-related points identified by *linear* symmetry transformations.



in depth:
appendix A25, p. 1048

11.2 Relative periodic orbits



So far we have demonstrated that symmetry relates classes of orbits. Now we show that a symmetry reduces computation of periodic orbits to repeats of shorter, ‘relative periodic orbit’ segments.

Equivariance of a flow under a symmetry means that the symmetry image of a cycle is again a cycle, with the same period and stability. The new orbit may be topologically distinct (in which case it contributes to the multiplicity (10.6) of the cycle) or it may be the same cycle.

A cycle p is G_p -*symmetric* under symmetry operation $g \in G_p$ if the operation acts on it as a shift in time, advancing a cycle point to a cycle point on the sym-

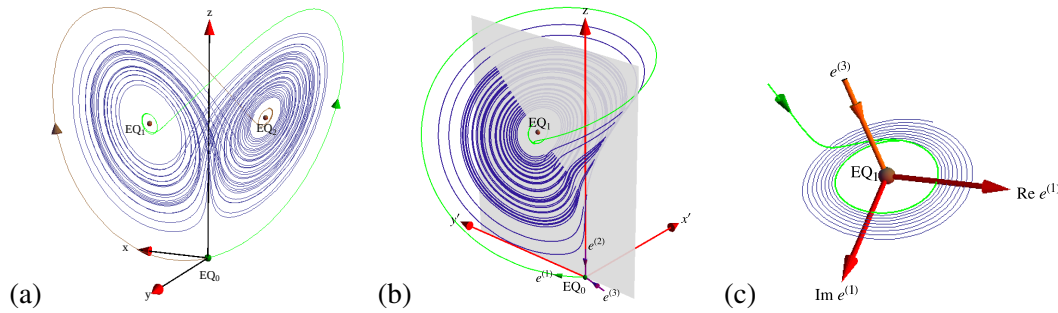


Figure 11.2: (a) Lorenz attractor of figure 3.5, the full state space coordinates $[x, y, z]$, with the unstable manifold orbits $W^u(EQ_0)$. (Green) is a continuation of the unstable $e^{(1)}$ of EQ_0 , and (brown) is its π -rotated symmetric partner. (b) Lorenz attractor plotted in $[\hat{x}, \hat{y}, z]$, the doubled-polar angle coordinates (11.13), with points related by π -rotation in the $[x, y]$ plane identified. Stable eigenvectors of EQ_0 : $e^{(3)}$ and $e^{(2)}$, along the z axis (11.12). Unstable manifold orbit $W^u(EQ_0)$ (green) is a continuation of the unstable $e^{(1)}$ of EQ_0 . (c) Blow-up of the region near EQ_1 : The unstable eigenplane of EQ_1 defined by $\text{Re } e^{(2)}$ and $\text{Im } e^{(2)}$, the stable eigenvector $e^{(3)}$. The descent of the EQ_0 unstable manifold (green) defines the innermost edge of the strange attractor. As it is clear from (a), it also defines its outermost edge. Work through examples 11.8 and 11.9. (E. Siminos)

metry related segment. The cycle p can thus be subdivided into m_p repeats of a *relative periodic orbit segment*, ‘prime’ in the sense that the full state space cycle is built from its repeats: see figure 11.1 for examples. Thus, in the presence of a discrete symmetry, the notion of a periodic orbit is replaced by the notion of the shortest segment of the full state space cycle, which tiles the cycle under the action of the group. In what follows, we refer to this segment as a *relative periodic orbit*. In the literature this is sometimes referred to as a *short periodic orbit*, or, for finite symmetry groups, as a *pre-periodic orbit*.

The relative periodic orbit p (or its *equivariant periodic orbit*) is the orbit $x(t)$ in state space $\mathcal{M}$ which exactly recurs

$$x(t) = g_p x(t + T_p) \quad (11.3)$$

for the shortest fixed *relative period* T_p and a fixed group action $g \in G_p$. These group actions are referred to as ‘shifts’ or, in the case of continuous symmetries, as ‘phases.’ For a discrete group $g^m = e$ and finite m (10.3), the period of the corresponding full state space orbit is given by the $m_p \times$ (period of the relative periodic orbit), $T_p = |G_p| T_{\bar{p}}$. Additionally, the i th Floquet multiplier $\Lambda_{p,i}$ is given by $\Lambda_{\bar{p},i}^{m_p}$ of the relative periodic orbit. The elements of the quotient space $b \in G/G_p$ generate the copies bp , so the multiplicity (10.6) of the full state space cycle p is $m_p = |G|/|G_p|$.

 example 11.10
p. 207

11.3 Dynamics reduced to fundamental domain

I submit my total lack of apprehension of fundamental concepts.

—John F. Gibson



So far we have used symmetry to effect a reduction in the number of independent cycles, by separating them into classes, and slicing them into ‘prime’ relative orbit segments. The next step achieves much more; it replaces each class by a single (typically shorter) prime cycle segment.

1. Discrete symmetry tessellates the state space into dynamically equivalent domains, and thus induces a natural partition of state space. If the dynamics is invariant under a discrete symmetry, the state space $\mathcal{M}$ can be completely tiled by a *fundamental domain* $\tilde{\mathcal{M}}$ and its symmetry images $\tilde{\mathcal{M}}_a = a\tilde{\mathcal{M}}$, $\tilde{\mathcal{M}}_b = b\tilde{\mathcal{M}}$, $\dots$ under the action of the symmetry group $G = \{e, a, b, \dots\}$,

$$\mathcal{M} = \tilde{\mathcal{M}} \cup \tilde{\mathcal{M}}_a \cup \tilde{\mathcal{M}}_b \cdots \cup \tilde{\mathcal{M}}_{|G|}. \quad (11.4)$$

See figure 10.1 for an example, which discusses the tiling of three disks with 6 copies of the fundamental domain.



2. Discrete symmetry can be used to restrict all computations to the fundamental domain $\tilde{\mathcal{M}} = \mathcal{M}/G$, the reduced state space quotient of the full state space $\mathcal{M}$ by the group actions of G . Several examples are given in figures 11.1, 11.5 and 11.3.

We can use the invariance condition (10.4) to move the starting point x into the fundamental domain $x = a\tilde{x}$, and then use the relation $a^{-1}b = h^{-1}$ to also relate the endpoint $y \in \tilde{\mathcal{M}}_b$ to its image in the fundamental domain $\tilde{\mathcal{M}}$. While the global trajectory runs over the full space $\mathcal{M}$, the restricted trajectory is brought back into the fundamental domain $\tilde{\mathcal{M}}$ any time it exits into an adjoining tile. The two trajectories are related by the symmetry operation h , which maps the global endpoint into its fundamental domain image.

3. Cycle multiplicities induced by the symmetry are removed by reduction of the full dynamics to the dynamics on a fundamental domain. Each symmetry-related set of global cycles p corresponds to precisely one fundamental domain (or relative) cycle $\tilde{p}$.
4. Conversely, each fundamental domain cycle $\tilde{p}$ traces out a segment of the global cycle p , with the end point of the cycle $\tilde{p}$ mapped into the irreducible segment of p with the group element $h_{\tilde{p}}$. A relative periodic orbit segment in the full state space is thus a periodic orbit in the fundamental domain.
5. The group elements $G = \{e, g_2, \dots, g_{|G|}\}$ which map the fundamental domain $\tilde{\mathcal{M}}$ into its copies $g\tilde{\mathcal{M}}$, serve also as letters of a symbolic dynamics alphabet.

exercise 11.3

For a symmetry reduction in presence of continuous symmetries, see sect. 13.2.

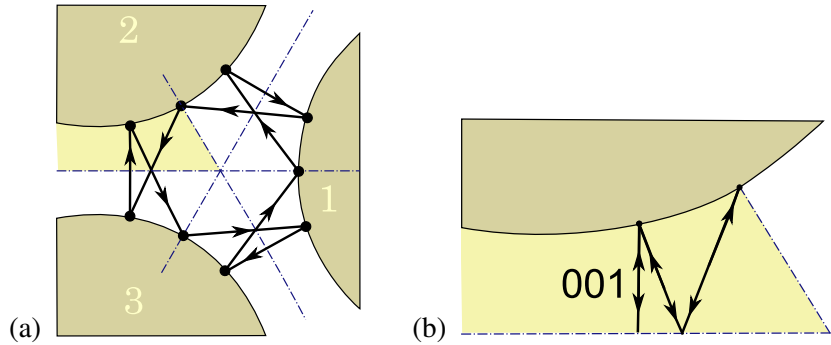


example 11.4
p. 204



example 11.11
p. 207

Figure 11.3: (a) The pair of full-space 9-cycles, the counter-clockwise $\overline{121232313}$ and the clockwise $\overline{131323212}$ correspond to (b) one fundamental domain 3-cycle $00\overline{1}$.



11.4 Life on the border

In what follows, we shall have to worry also about the boundaries that define the fundamental domain $\tilde{\mathcal{M}}$. Here we by definition include boundaries into the fundamental tile. The state space transformation $g \in G$ leaves invariant the set of *boundary points* (see (11.4))

$$\mathcal{M}_B = \tilde{\mathcal{M}} \cap \tilde{\mathcal{M}}_a \cap \tilde{\mathcal{M}}_b \cdots \cap \tilde{\mathcal{M}}_{|G|}. \quad (11.5)$$

Peculiar effects, however, arise for orbits that run along symmetry lines that border a fundamental domain. For example, under reflection σ across a symmetry axis, the axis itself remains invariant. The properties of boundary orbits that belong to G -fixed (point-wise invariant) boundary sets will require a bit of thinking. In our 3-disk example, no such orbits are possible, but they exist in other systems, such as the bounded region of the Hénon-Heiles potential (remark 11.2), in $1d$ maps of example 11.3, and in Lorenz flow of example 11.8, where the z axis is a G -invariant border. While boundary orbits are invariant under some symmetry operations, their neighborhoods are not.

That's why one sometimes surgically removes boundaries and defines free action.

Definition: Free action. An group action on a state space submanifold $\hat{\mathcal{M}}$ is free if all of the isotropy subgroups $G_x, x \in \hat{\mathcal{M}}$ are trivial.

The fact that open neighborhoods of the border are in part outside of it complicates analysis (linear stability of orbits within the boundary has eigenvectors in the full state space). This affects the Jacobian matrix M_p of the orbit and its Floquet multipliers.

Typically, for low-dimensional state spaces, relatively few boundary orbits exist, and they tend to be among the shortest orbits. Thus, they play a key role in dynamics.



section 25.4.3

11.5 Invariant polynomials

All invariants are expressible in terms of a finite number among them. We cannot claim its validity for every group G ; rather, it will be our chief task to investigate for each particular group whether a finite integrity basis exists or not; the answer, to be sure, will turn out affirmative in the most important cases.

—Hermann Weyl, a motivational quote on the “so-called first main theorem of invariant theory”



Physical laws should have the same form in symmetry-equivalent coordinate frames, so they are often formulated in terms of functions (Hamiltonians, Lagrangians, ...) invariant under a given set of symmetries.

Definition: G -invariant function. A function is said to be G -invariant if

$$f(gx) = f(x), x \in \mathcal{M}. \quad (11.6)$$

A G -invariant function is constant along the group orbit of x .

Invariant polynomial functions play a particularly important role in invariant theory. The set of all G -invariant polynomial functions of x which is finitely generated, according to the key result of the representation theory of invariant functions is:

Hilbert-Weyl theorem. For a compact group G , there exists a finite G -invariant homogenous polynomial basis $\{u_1, u_2, \dots, u_m\}$, $m \geq d$, such that any G -invariant polynomial can be written as a multinomial

$$h(x) = p(u_1(x), u_2(x), \dots, u_m(x)), \quad x \in \mathcal{M}. \quad (11.7)$$

These polynomials are linearly independent, but can be functionally dependent through nonlinear relations called *syzygies*.

In practice, explicit construction of G -invariant basis can be a laborious undertaking, and we will not take this path except for a few simple low-dimensional cases, such as ‘doubled-polar angle representation’ (11.13) and the 5-dimensional example of sect. 13.6. We prefer to apply the symmetry to the system as given, rather than undertake a series of nonlinear coordinate transformations that the theorem suggests. (What ‘compact’ in the above refers to will become clearer after we have discussed continuous symmetries. For now, it suffices to know that any finite discrete group is compact.)

exercise 11.2



example 11.12
p. 208

Résumé

If a dynamical system $(\mathcal{M}, f)$ has a symmetry G , the symmetry should be deployed to ‘quotient’ the state space to fundamental domain $\hat{\mathcal{M}} = \mathcal{M}/G$, i.e., identify all symmetry-equivalent $x \in \mathcal{M}$ on each group orbit, thus replacing the full state space dynamical system $(\mathcal{M}, f)$ by the symmetry-reduced $(\hat{\mathcal{M}}, \hat{f})$. The main result of this chapter can be stated as follows:

In the presence of a discrete symmetry G , associated with each full state space solution p is the group of its symmetries $G_p \subseteq G$ of order $1 \leq |G_p| \leq |G|$, whose elements leave the orbit $\mathcal{M}_p$ invariant. The elements of G_p act on $\mathcal{M}_p$ as shifts, tiling it with $|G_p|$ copies of its shortest invariant segment, the relative periodic orbit $\tilde{p}$. The elements of the coset $b \in G/G_p$ generate $m_p = |G|/|G_p|$ equivalent copies of p .

Once you grasp the relation between the full state space $\mathcal{M}$ and the desymmetrized, G -quotiented reduced state space (fundamental domain) $\mathcal{M}/G$, you will find the life as a fundamentalist so much simpler that you will never return to your full state space ways of yesteryear. The reduction to the fundamental domain $\tilde{\mathcal{M}} = \mathcal{M}/G$ simplifies symbolic dynamics and eliminates symmetry-induced degeneracies. For short orbits, the labor savings is dramatic. For example, 256 periodic points of period 8 exist for the 3-disk game of pinball, but reduction to fundamental domain, non-degenerate prime cycles reduces this number to 30. By chapter 24, the savings will be even more dramatic; relative periodic orbits will tile the infinite periodic state space, and replace a numerical simulation of diffusion in the infinite domain by an exact calculation of the diffusion constant, on a compact torus.

Commentary

Remark 11.1. Symmetries of the Lorenz equation. (Continued from remark 2.3)

After having studied example 11.8 you will appreciate why ChaosBook.org starts out with the symmetry-less Rössler flow (2.30), instead of the better known Lorenz flow (2.25). Indeed, getting rid of symmetry was one of Rössler’s motivations. He threw the baby out with the water; for Lorenz flow dimensionalities of stable/unstable manifolds make possible a robust heteroclinic connection absent from Rössler flow, with unstable manifold of an equilibrium flowing into the stable manifold of another equilibrium. How such connections are forced upon us is best grasped by perusing chapter 13 ‘Heteroclinic tangles’ of the inimitable Abraham and Shaw Classics Illustrated [1]. Their beautiful hand-drawn sketches elucidate the origin of heteroclinic connections in the Lorenz flow (and its high-dimensional Navier-Stokes relatives) better than any computer simulation. Miranda and Stone [23] were the first to quotient the C_2 symmetry and explicitly construct the desymmetrized, ‘proto-Lorenz system’, by a nonlinear coordinate transformation into the Hilbert-Weyl polynomial basis invariant under the action of the symmetry group [6]. For in-depth discussion of symmetry-reduced (‘images’) and symmetry-extended (‘covers’) topology, symbolic dynamics, periodic orbits, invariant polynomial bases etc., of

Lorenz, Rössler and many other low-dimensional systems, there is no better reference than the Gilmore and Letellier monograph [11]. They interpret [19] the proto-Lorenz and its ‘double cover’ Lorenz as ‘intensities’ being the squares of ‘amplitudes’, and call quotiented flows such as (Lorenz)/ C_2 ‘images.’ Our ‘doubled-polar angle’ visualization of figure 14.14 is a proto-Lorenz in disguise; we, however, integrate the flow and construct Poincaré sections and return maps in the original Lorenz $[x, y, z]$ coordinates, without any nonlinear coordinate transformations. The return map figure 14.15 is reminiscent in shape both of the one given by Lorenz in his original paper, and the one plotted in a radial coordinate by Gilmore and Letellier. Nevertheless, it is profoundly different; our return maps are from unstable manifold $\rightarrow$ itself, and thus intrinsic and coordinate independent. In this we follow Christiansen *et al.* [5]. This construction is necessary for high-dimensional flows in order to avoid problems such as double-valuedness of return map projections on arbitrary 1-dimensional coordinates, encountered already in the Rössler example of figure 3.4. More importantly, as we know the embedding of the unstable manifold into the full state space, a periodic point of our return map *is* - regardless of the period of the cycle - the periodic point in the full state space, so no additional Newton searches are needed. In homage to Lorenz, we note that his return map was already symmetry-reduced; as z belongs to the symmetry invariant $\text{Fix}(G)$ subspace, one can replace dynamics in the full space by $\dot{z}, \ddot{z}, \dots$. That is G -invariant by construction [11].

Remark 11.2. Examples of systems with discrete symmetries. Almost any flow of interest is symmetric in some way or other. The list of examples is endless, and here we list a handful that we find interesting: C_2 symmetry in the Lorenz system (remark 2.3), Ising model, 3-dimensional anisotropic Kepler potential [4, 12, 25], D_4 symmetry in quartic oscillators [8, 21], the pure x^2y^2 potential [3, 22], hydrogen in a magnetic field [9], and D_2 symmetry in the stadium billiard [24]. A number of nontrivial desymmetrizations are carried out in the Balász and Voros review [2]. An example of a system with D_3 symmetry is provided by the motion of a particle in the Hénon-Heiles potential [13–15, 17], as well as in the Chernoff-Barrow-Lifshitz-Khalatnikov-Sinai-Khanin-Shchur cosmology [18].

$$V(r, \theta) = \frac{1}{2}r^2 + \frac{1}{3}r^3 \sin(3\theta) .$$

Our 3-disk coding is insufficient for this system because of the existence of elliptic islands and because the three orbits that run along the symmetry axis cannot be labeled in our code. As these orbits run along the boundary of the fundamental domain, they require the special treatment. A partial classification of the 67 possible symmetries of solutions of the plane Couette flow of example 10.10, and their reduction to 5 conjugate classes is given in ref. [10].

Remark 11.3. Index. For a finite group G , the index (10.6) is the number $|G|/|G_p|$ of left cosets of a subgroup G_p . In our applications the time orbit p is invariant under all actions of its symmetry group G_p , and each coset corresponds to a conjugate symmetry subgroup (10.7), so on page 193 we refer to the index $m_p = |G|/|G_p|$ as the ‘multiplicity’ of orbit p . For infinite discrete groups, such as the space groups of chapter 24, the index is finite, even though both a subgroup and the number of its cosets are infinite, see Wikipedia [26] [Index of a subgroup](#), p. 90 of Dummit and Foote [7], and refs. [16, 20].

Remark 11.4. Time reversal. Some background on D_∞ time-reversal symmetry of discrete time dynamics [20]: Time-reversed orbits not self-dual under time reversal are here counted as pairs, see example 15.6, figure 15.14, table 18.1, table 34.1, table 18.4, table 18.5, table 15.1, table 15.2, table A25.1, example 15.3, example 30.6, example A10.1, exercise 7.2, exercise 17.1, exercise 23.8, exercise 25.4, exercise 25.5, exercise 2.2, sect. A10.3, the time-reversal discussion in sect. 42.2.3 and figure 42.5.

References

- [1] R. H. Abraham and C. D. Shaw, *Dynamics - The Geometry of Behavior* (Wesley, Reading, MA, 1992).
- [2] N. L. Balász and A. Voros, “Chaos on the pseudosphere”, *Phys. Rep.* **143**, 109–240 (1986).
- [3] A. Carnegie and I. C. Percival, “Regular and chaotic motion in some quartic potentials”, *J. Phys. A* **17**, 801 (1984).
- [4] F. Christiansen and P. Cvitanović, “Periodic orbit quantization of the anisotropic Kepler problem”, *Chaos* **2**, 61–69 (1992).
- [5] F. Christiansen, P. Cvitanović, and V. Putkaradze, “Hopf’s last hope: Spatiotemporal chaos in terms of unstable recurrent patterns”, *Nonlinearity* **10**, 55–70 (1997).
- [6] D. A. Cox, J. B. Little, and D. O’Shea, *Ideals, Varieties and Algorithms* (Springer, New York, 2007).
- [7] D. S. Dummit and R. M. Foote, *Abstract Algebra* (Wiley, 2003).
- [8] B. Eckhardt, G. Hose, and E. Pollak, “Quantum mechanics of a classically chaotic system: Observations on scars, periodic orbits, and vibrational adiabaticity”, *Phys. Rev. A* **39**, 3776–3793 (1989).
- [9] B. Eckhardt and D. Wintgen, “Symbolic description of periodic orbits for the quadratic Zeeman effect”, *J. Phys. B* **23**, 355–363 (1990).
- [10] J. F. Gibson, J. Halcrow, and P. Cvitanović, “Equilibrium and traveling-wave solutions of plane Couette flow”, *J. Fluid Mech.* **638**, 243–266 (2009).
- [11] R. Gilmore and C. Letellier, *The Symmetry of Chaos* (Oxford Univ. Press, Oxford, 2007).
- [12] M. C. Gutzwiller, “The quantization of a classically ergodic system”, *Physica D* **5**, 183–207 (1982).
- [13] M. Hénon and C. Heiles, “The applicability of the third integral of motion: Some numerical experiments”, *Astron. J.* **69**, 73–79 (1964).
- [14] C. Jung and P. Richter, “Classical chaotic scattering-periodic orbits, symmetries, multifractal invariant sets”, *J. Phys. A* **23**, 2847 (1990).
- [15] C. Jung and H. J. Scholz, “Cantor set structures in the singularities of classical potential scattering”, *J. Phys. A* **20**, 3607 (1987).
- [16] Y.-O. Kim, J. Lee, and K. K. Park, “A zeta function for flip systems”, *Pacific J. Math.* **209**, 289–301 (2003).
- [17] B. Lauritzen, “Discrete symmetries and the periodic-orbit expansions”, *Phys. Rev. A* **43**, 603–606 (1991).
- [18] O. M. Lecian, “Reflections on the hyperbolic plane”, *Int. J. Mod. Phys. D* **22**, 1350085 (2013).
- [19] C. Letellier and R. Gilmore, “Covering dynamical systems: Two-fold covers”, *Phys. Rev. E* **63**, 016206 (2001).

- [20] H. Liang and P. Cvitanović, “A chaotic lattice field theory in one dimension”, *J. Phys. A* **55**, 304002 (2022).
- [21] C. C. Martens, R. L. Waterland, and W. P. Reinhardt, “Classical, semiclassical, and quantum mechanics of a globally chaotic system: Integrability in the adiabatic approximation”, *J. Chem. Phys.* **90**, 2328 (1989).
- [22] S. G. Matanyan, G. K. Savvidy, and N. G. Ter-Arutyunyan-Savvidy, “Classical Yang-Mills mechanics. Nonlinear colour oscillations”, *Sov. Phys. JETP* **80**, 830–838 (1981).
- [23] R. Miranda and E. Stone, “The proto-Lorenz system”, *Phys. Lett. A* **178**, 105–113 (1993).
- [24] J. M. Robbins, “Discrete symmetries in periodic-orbit theory”, *Phys. Rev. A* **40**, 2128–2136 (1989).
- [25] G. Tanner and D. Wintgen, “Quantization of chaotic systems”, *Chaos* **2**, 53 (1992).
- [26] Wikipedia contributors, [Index of a subgroup — Wikipedia, The Free Encyclopedia](#), 2022.

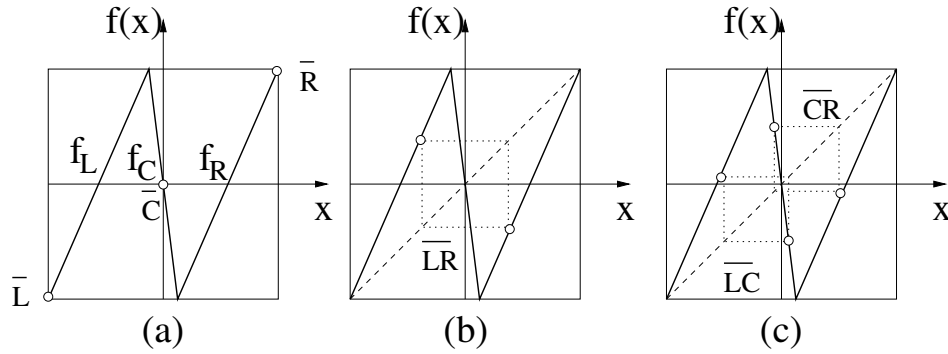


Figure 11.4: The D_1 -equivariant bimodal sawtooth map of figure 10.2 has three types of periodic orbits: (a) D_1 -fixed fixed point $\bar{C}$, asymmetric fixed points pair $\{\bar{L}, \bar{R}\}$. (b) D_1 -symmetric (setwise invariant) 2-cycle $\bar{L}\bar{R}$, composed of the relative cycle segment from L to R and its repeat from R to L . (c) Asymmetric 2-cycles pair $\{\bar{L}\bar{C}, \bar{C}\bar{R}\}$. (study example 11.3; continued in figure 11.5) (Y. Lan)

11.6 Examples

Example 11.1. D_1 -symmetric cycles. (Continued from example 10.5) For D_1 the period of a set-wise symmetric cycle is even ($n_s = 2n_{\tilde{s}}$), and the mirror image of the x_s periodic point is reached by traversing the relative periodic orbit segment $\tilde{s}$ of length $n_{\tilde{s}}$, $f^{n_{\tilde{s}}}(x_s) = sx_s$, see figure 11.4 (b).

[return to p. 193](#)

Example 11.2. D_1 -invariant cycles. In the example at hand there is only one G -invariant (point-wise invariant) orbit, the fixed point $\bar{C}$ at the origin, see figure 11.4 (a). As reflection symmetry is the only discrete symmetry that a map of the interval can have, this example completes the group-theoretic analysis of 1-dimensional maps. We shall continue analysis of this system in example 11.4, and work out the symbolic dynamics of such reflection symmetric systems in example 15.6.

[return to p. 194](#)

Example 11.3. Group D_1 - a reflection symmetric 1d map. Consider the bimodal ‘sawtooth’ map of figure 11.4, with the state space $\mathcal{M} = [-1, 1]$ split into three regions $\mathcal{M} = \{\mathcal{M}_L, \mathcal{M}_C, \mathcal{M}_R\}$ which we label with a 3-letter alphabet L (eft), C (enter), and R (ight). The symbolic dynamics is complete ternary dynamics, with any sequence of letters $\mathcal{A} = \{L, C, R\}$ corresponding to an admissible trajectory (‘complete’ means no additional grammar rules required, see example 14.7 below). The D_1 -equivariance of the map, $D_1 = \{e, s\}$, implies that if $\{x_n\}$ is a trajectory, so is $\{sx_n\}$.

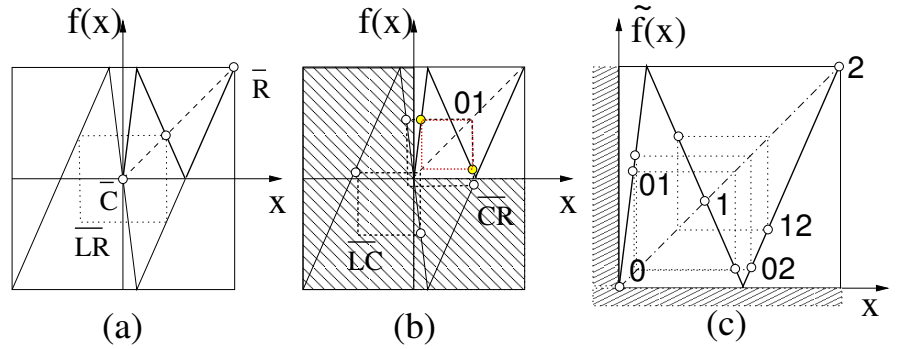


$\text{Fix}(G)$, the set of points invariant under group action of D_1 , $\tilde{\mathcal{M}} \cap s\tilde{\mathcal{M}}$, is just this fixed point $x = 0$, the reflection symmetry point. If a is an asymmetric cycle, s maps it into the reflected cycle sa , with the same period and the same stability properties, see the fixed points pair $\{\bar{L}, \bar{R}\}$ and the 2-cycles pair $\{\bar{L}\bar{C}, \bar{C}\bar{R}\}$ in figure 11.4 (c).

[return to p. 194](#)

Example 11.4. Group D_1 and reduction to the fundamental domain. Consider again the reflection-symmetric bimodal Ulam sawtooth map $f(-x) = -f(x)$ of example 11.3, with symmetry group $D_1 = \{e, s\}$. The state space $\mathcal{M} = [-1, 1]$ can be tiled by half-line

Figure 11.5: The bimodal Ulam sawtooth map of figure 11.4 with the D_1 symmetry $f(-x) = -f(x)$, restricted to the fundamental domain. $f(x)$ is indicated by the thin line, and fundamental domain map $\tilde{f}(\tilde{x})$ by the thick line. (a) Boundary fixed point $\bar{C}$ is the fixed point $\bar{0}$. The asymmetric fixed point pair $\{\bar{L}, \bar{R}\}$ is reduced to the fixed point $\bar{2}$, and the full state space symmetric 2-cycle $\bar{LR}$ is reduced to the fixed point $\bar{1}$. (b) The asymmetric 2-cycle pair $\{\bar{LC}, \bar{CR}\}$ is reduced to 2-cycle $0\bar{1}$. (c) All fundamental domain fixed points and 2-cycles. (work through example 11.4) (Y. Lan)



$\tilde{\mathcal{M}} = [0, 1]$, and $s\tilde{\mathcal{M}} = [-1, 0]$, its image under a reflection across $x = 0$ point. The dynamics can then be restricted to the *fundamental domain* $\tilde{x}_k \in \tilde{\mathcal{M}} = [0, 1]$; every time a trajectory leaves this interval, it is mapped back using s .

In figure 11.5 the fundamental domain map $\tilde{f}(\tilde{x})$ is obtained by reflecting $x < 0$ segments of the global map $f(x)$ into the upper right quadrant. $\tilde{f}$ is also bimodal and piecewise-linear, with $\tilde{\mathcal{M}} = [0, 1]$ split into three regions $\tilde{\mathcal{M}} = \{\tilde{\mathcal{M}}_0, \tilde{\mathcal{M}}_1, \tilde{\mathcal{M}}_2\}$ which we label with a 3-letter alphabet $\tilde{\mathcal{A}} = \{0, 1, 2\}$. The symbolic dynamics is again complete ternary dynamics, with any sequence of letters $\{0, 1, 2\}$ admissible.

However, the interpretation of the ‘desymmetrized’ dynamics is quite different - the multiplicity of every periodic orbit is now 1, and relative periodic segments of the full state space dynamics are all periodic orbits in the fundamental domain. Consider figure 11.5:

In (a) the boundary fixed point $\bar{C}$ is also the fixed point $\bar{0}$. The asymmetric fixed point pair $\{\bar{L}, \bar{R}\}$ is reduced to the fixed point $\bar{2}$, and the full state space symmetric 2-cycle $\bar{LR}$ is reduced to the fixed point $\bar{1}$. (b) The asymmetric 2-cycle pair $\{\bar{LC}, \bar{CR}\}$ is reduced to the 2-cycle $0\bar{1}$. Finally, the symmetric 4-cycle $\bar{LCRC}$ is reduced to the 2-cycle $0\bar{2}$. This completes the conversion from the full state space for all fundamental domain fixed points and 2-cycles, frame (c).

[return to p. 197](#)

Example 11.5. D_3 symmetry of the 3-disk game of pinball. If the three unit-radius disks in figure 10.1 are equidistantly spaced, our game of pinball has a sixfold symmetry. The symmetry group of relabeling the 3 disks is the permutation group S_3 ; however, it is more instructive to think of this group geometrically, as the dihedral group D_3

$$D_3 = \{e, r, r^2, s, sr, sr^2\}, \quad (11.8)$$

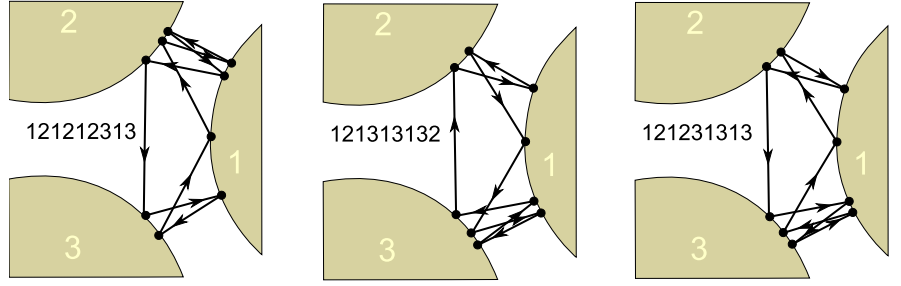
the group of order $|G| = 6$ consisting of the identity element e , three reflections across symmetry axes $\{s, sr, sr^2\}$, and two rotations by $2\pi/3$ and $4\pi/3$ denoted $\{r, r^2\}$. (continued in example 11.6)

[return to p. 194](#)

Example 11.6. 3-disk game of pinball - symmetry-related orbits. (Continued from example 11.5) Applying an element (identity, rotation by $\pm 2\pi/3$, or one of the three possible reflections) of this symmetry group to a trajectory yields another trajectory. For instance, s_2 , the flip across the symmetry axis going through disk 1 interchanges the symbols 2 and 3; it maps the cycle $\bar{1}2\bar{1}2\bar{3}$ into $\bar{1}3\bar{1}3\bar{2}$, figure 11.1 (c). Cycles $\bar{1}2$, $\bar{2}3$, and $\bar{1}3$ in figure 11.1 (a) are related to each other by rotation by $\pm 2\pi/3$, or, equivalently, by a relabeling of the disks. (continued in example 10.8)

[return to p. 194](#)

Figure 11.6: Cycle $\overline{121212313}$ has multiplicity 6; shown here is $\overline{121313132} = s_2 \overline{121212313}$. However, $\overline{121231313}$ which has the same stability and period is related to $\overline{121313132}$ by time reversal, but not by any spatial D_3 symmetry.



Example 11.7. 3-disk game of pinball - cycle symmetries.

(Continued from

example 10.9) The C_3 subgroup $G_p = \{e, r, r_2\}$ invariance is exemplified by the two cycles $\overline{123}$ and $\overline{132}$ which are invariant under rotations by $2\pi/3$ and $4\pi/3$, but are mapped into each other by any reflection, figure 11.6 (a), and have multiplicity (10.6) $|G|/|G_p| = 2$.

The D_1 type of a subgroup is exemplified by the symmetries of $\hat{p} = 1213$. This cycle is invariant under reflection $s_2\{1213\} = \overline{1312} = \overline{1213}$, so the invariant subgroup is $G_{\hat{p}} = \{e, s_2\}$, with multiplicity $m_{\hat{p}} = |G|/|G_p| = 3$; the cycles in this class, $\overline{1213}$, $\overline{1232}$ and $\overline{1323}$, are related by $2\pi/3$ rotations, figure 11.6 (b).

A cycle of no symmetry, such as $\overline{12323}$, has $G_p = \{e\}$ and contributes in all six copies (the remaining cycles in the class are $\overline{12132}$, $\overline{12313}$, $\overline{12323}$, $\overline{13132}$ and $\overline{13232}$), figure 11.6 (c).

Besides the spatial symmetries, for Hamiltonian systems cycles may be related by time reversal symmetry. An example are the cycles $\overline{121212313}$ and $\overline{313212121} = \overline{121213132}$ which have the same periods and stabilities, but are related by no space symmetry, see figure 11.6. (continued in example 11.11)

[return to p. 194](#)

Example 11.8. Desymmetrization of Lorenz flow.

(Continuation of exam-

ple 10.6) Lorenz equation (10.15) is equivariant under (10.16), the action of order-2 group $C_2 = \{e, r\}$, where r is $[x, y]$ -plane, half-cycle rotation by π about the z -axis:

$$(x, y, z) \rightarrow r(x, y, z) = (-x, -y, z). \quad (11.9)$$

$(r)^2 = 1$ condition decomposes the state space into two linearly irreducible subspaces $\mathcal{M} = \mathcal{M}^+ \oplus \mathcal{M}^-$, the z -axis $\mathcal{M}^+$ and the $[x, y]$ plane $\mathcal{M}^-$, with projection operators onto the two subspaces given by

$$P^+ = \frac{1}{2}(1 + r) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad P^- = \frac{1}{2}(1 - r) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (11.10)$$

As the flow is C_2 -invariant, so is its linearization $\dot{x} = Ax$. Evaluated at EQ_0 , A commutes with r , and, as we have already seen in example 4.6, the EQ_0 stability matrix decomposes into $[x, y]$ and z blocks.

The 1-dimensional $\mathcal{M}^+$ subspace is the fixed-point subspace, with the z -axis points left *point-wise invariant* under the group action

$$\mathcal{M}^+ = \text{Fix}(C_2) = \{x \in \mathcal{M} \mid g x = x \text{ for } g \in \{e, r\}\} \quad (11.11)$$

(here $x = (x, y, z)$ is a 3-dimensional vector, not the coordinate x). A C_2 -fixed point $x(t)$ in $\text{Fix}(C_2)$ moves with time, but according to (10.11) remains within $x(t) \in \text{Fix}(C_2)$ for all

times; the subspace $\mathcal{M}^+ = \text{Fix}(C_2)$ is *flow invariant*. In case at hand this jargon is a bit of an overkill: clearly for $(x, y, z) = (0, 0, z)$ the full state space Lorenz equation (10.15) is reduced to the exponential contraction to the EQ_0 equilibrium,

$$\dot{z} = -bz. \quad (11.12)$$

However, for higher-dimensional flows the flow-invariant subspaces can be high-dimensional, with interesting dynamics of their own. Even in this simple case this subspace plays an important role as a topological obstruction: the orbits can neither enter it nor exit it, so the number of windings of a trajectory around it provides a natural, topological symbolic dynamics.

The $\mathcal{M}^-$ subspace is, however, *not* flow-invariant, as the nonlinear terms $\dot{z} = xy - bz$ in the Lorenz equation (10.15) send all initial conditions within $\mathcal{M}^- = (x(0), y(0), 0)$ into the full, $z(t) \neq 0$ state space $\mathcal{M}/\mathcal{M}^+$. (continued in example 11.9)

[return to p. 195](#)

(E. Siminos and J. Halcrow)

Example 11.9. Lorenz flow in doubled-polar angle representation. By taking as a Poincaré section any r -equivariant, non-self-intersecting surface that contains the z axis, the state space is divided into a half-space fundamental domain $\tilde{\mathcal{M}} = \mathcal{M}/C_2$ and its 180° rotation $r\tilde{\mathcal{M}}$. An example is afforded by the $\mathcal{P}$ plane section of the Lorenz flow in figure 3.5. Take the fundamental domain $\tilde{\mathcal{M}}$ to be the half-space between the viewer and $\mathcal{P}$. Then the full Lorenz flow is captured by re-injecting back into $\tilde{\mathcal{M}}$ any trajectory that exits it, by a rotation of π around the z axis.

[chapter 30](#)

As any such r -invariant section does the job, a choice of a ‘fundamental domain’ is here largely matter of taste. For purposes of visualization it is convenient to make the double-cover nature of the full state space by $\tilde{\mathcal{M}}$ explicit, through any state space redefinition that maps a pair of points related by symmetry into a single point. In case at hand, this can be easily accomplished by expressing (x, y) in polar coordinates $(x, y) = (R \cos \theta, R \sin \theta)$, and then plotting the flow in the ‘*doubled-polar angle representation*.’

[section 11.5](#)
[exercise 11.4](#)

$$(\hat{x}, \hat{y}, z) = (R \cos 2\theta, R \sin 2\theta, z) = ((x^2 - y^2)/r, 2xy/r, z), \quad (11.13)$$

as in figure 11.2(b). In contrast to the original G -equivariant coordinates $[x, y, z]$, the Lorenz flow expressed in the new coordinates $[\hat{x}, \hat{y}, z]$ is G -invariant, see example 11.12. In this representation the $\tilde{\mathcal{M}} = \mathcal{M}/C_2$ fundamental domain flow is a smooth, continuous flow, with (any choice of) the fundamental domain stretched out to seamlessly cover the entire $[\hat{x}, \hat{y}]$ plane. (continued in example 14.4)

[return to p. 195](#)

(E. Siminos and J. Halcrow)

Example 11.10. Relative periodic orbits of Lorenz flow. (Continuation of example 11.8) The relation between the full state space periodic orbits, and the fundamental domain (11.13) reduced relative periodic orbits of the Lorenz flow: an asymmetric full state space cycle pair p, rp maps into a single cycle $\tilde{p}$ in the fundamental domain, and any self-dual cycle $p = rp = \tilde{p}r\tilde{p}$ is a repeat of a relative periodic orbit $\tilde{p}$.

[return to p. 196](#)

Example 11.11. 3-disk game of pinball in the fundamental domain.

If the dynamics is equivariant under interchanges of disks, the absolute disk labels $\epsilon_i = 1, 2, \dots, N$ can be replaced by the symmetry-invariant relative disk→disk increments g_i , where g_i is the discrete group element that maps disk $i-1$ into disk i . For 3-disk system

g_i is either reflection σ back to initial disk (symbol '0') or $2\pi/3$ rotation by C to the next disk (symbol '1'). An immediate gain arising from symmetry invariant relabeling is that N -disk symbolic dynamics becomes $(N-1)$ -nary, with no restrictions on the admissible sequences.

An irreducible segment corresponds to a periodic orbit in the *fundamental domain*, a one-sixth slice of the full 3-disk system, with the symmetry axes acting as reflecting mirrors (see figure 11.1(d)). A set of orbits (10.6) related in the full space by discrete symmetries maps onto a single fundamental domain orbit. The reduction to the fundamental domain desymmetrizes the dynamics and removes all global discrete symmetry-induced degeneracies: rotationally symmetric global orbits (such as the 3-cycles $\overline{123}$ and $\overline{132}$) have multiplicity 2, reflection symmetric ones (such as the 2-cycles $\overline{12}$, $\overline{13}$ and $\overline{23}$) have multiplicity 3, and global orbits with no symmetry are 6-fold degenerate. Table 15.2 lists some of the shortest binary symbols strings, together with the corresponding full 3-disk symbol sequences and orbit symmetries. Some examples of such orbits are shown in figures 11.6 and 11.3. (continued in example 15.7)

[return to p. 197](#)

Example 11.12. Polynomials invariant under discrete operations on $\mathbb{R}^3$. (Continued from example 10.3) s is a *reflection* through the $[x, y]$ plane. Any $\{e, s\}$ -invariant function can be expressed in the polynomial basis $\{u_1, u_2, u_3\} = \{x, y, z^2\}$.

r is a $[x, y]$ -plane *rotation* by π about the z -axis. Any $\{e, r\}$ -invariant function can be expressed in the polynomial basis $\{u_1, u_2, u_3, u_4\} = \{x^2, xy, y^2, z\}$, with one syzygy between the basis polynomials, $(x^2)(y^2) - (xy)^2 = 0$.

P is an *inversion* through the point $(0, 0, 0)$. Any $\{e, P\}$ -invariant function can be expressed in the polynomial basis $\{u_1, \dots, u_6\} = \{x^2, y^2, z^2, xy, xz, yz\}$, with three syzygies between the basis polynomials, $(x^2)(y^2) - (xy)^2 = 0$, and its 2 permutations.

For the D_2 dihedral group $G = \{e, s, r, P\}$ the G -invariant polynomial basis is $\{u_1, u_2, u_3, u_4\} = \{x^2, y^2, z^2, xy\}$, with one syzygy, $(x^2)(y^2) - (xy)^2 = 0$.

[return to p. 199](#)

11.7 Exercises

11.1. $G_x \subset G$. The maximal set of group actions which maps a state space point x into itself,

$$G_x = \{g \in G : gx = x\}, \quad (11.14)$$

is called the *isotropy group* (or *stability subgroup* or *little group*) of x . Prove that the set G_x as defined in (11.14) is a subgroup of G .

11.2. **Polynomials invariant under discrete operations on $\mathbb{R}^3$.** Prove that the $\{e, s\}$, $\{e, r\}$, $\{e, P\}$ and $\{e, s, r, P\}$ -invariant polynomial basis and syzygies are those listed in example 11.12.

11.3. **Reduction of 3-disk symbolic dynamics to binary.** (continued from exercise 1.1)

- Verify that the 3-disk cycles $\overline{12}, \overline{13}, \overline{23}$, $\{123, 132\}$, $\{1213 + 2 \text{ perms.}\}$, $\{121232313 + 5 \text{ perms.}\}$, $\{121323 + 2 \text{ perms.}\}$, $\dots$, correspond to the fundamental domain cycles $\overline{0}, \overline{1}, \overline{01}, \overline{001}, \overline{011}, \dots$ respectively.
- Check the reduction for short cycles in table 15.2 by drawing them both in the full 3-disk system and in the fundamental domain, as in figure 11.3.
- Optional: Can you see how the group elements listed in table 15.2 relate irreducible segments to the fundamental domain periodic orbits?

(continued in exercise 15.7)

11.4. **Lorenz system in polar coordinates: group theory.** Use (A2.13), (A2.14) to rewrite the Lorenz equation (10.19) in polar coordinates (r, θ, z) , where $(x, y) = (r \cos \theta, r \sin \theta)$.

- Show that in the polar coordinates Lorenz flow takes form

$$\begin{aligned} \dot{r} &= \frac{r}{2} (-\sigma - 1 + (\sigma + \rho - z) \sin 2\theta \\ &\quad + (1 - \sigma) \cos 2\theta) \\ \dot{\theta} &= \frac{1}{2} (-\sigma + \rho - z + (\sigma - 1) \sin 2\theta \\ &\quad + (\sigma + \rho - z) \cos 2\theta) \\ \dot{z} &= -bz + \frac{r^2}{2} \sin 2\theta. \end{aligned} \quad (11.15)$$

- Argue that the transformation to polar coordinates is invertible almost everywhere. Where does the inverse not exist? What is group-theoretically special about the subspace on which the inverse not exist?
- Show that this is the $(\text{Lorenz})/Z_2$ quotient map for the Lorenz flow, i.e., that it identifies points related by the π rotation in the $[x, y]$ plane.
- Rewrite (10.19) in the invariant polynomial basis of example 11.12 and exercise 11.15.
- Show that a periodic orbit of the Lorenz flow in polar representation (11.15) is either a periodic orbit or a relative periodic orbit (11.3) of the Lorenz flow in the (x, y, z) representation.

By going to polar coordinates we have quotiented out the π -rotation $(x, y, z) \rightarrow (-x, -y, z)$ symmetry of the Lorenz equations, and constructed an explicit representation of the desymmetrized Lorenz flow.

11.5. **Proto-Lorenz system.** Here we quotient out the Z_2 symmetry by constructing an explicit “intensity” representation of the desymmetrized Lorenz flow, following Miranda and Stone [23].

1. Rewrite the Lorenz equation (10.15) in terms of variables

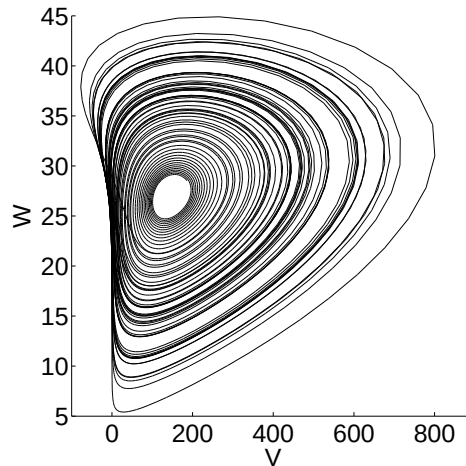
$$(u, v, z) = (x^2 - y^2, 2xy, z), \quad (11.16)$$

show that it takes form

$$\begin{bmatrix} \dot{u} \\ \dot{v} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} -(\sigma + 1)u + (\sigma - r)v + (1 - \sigma)N + vz \\ (r - \sigma)u - (\sigma + 1)v + (r + \sigma)N - uz - uN \\ v/2 - bz \end{bmatrix}$$

$$N = \sqrt{u^2 + v^2}. \quad (11.17)$$

2. Show that this is the (Lorenz)/ Z_2 quotient map for the Lorenz flow, i.e., that it identifies points related by the π rotation (11.9).
3. Show that (11.16) is invertible. Where does the inverse not exist?
4. Compute the equilibria of proto-Lorenz and their stabilities. Compare with the equilibria of the Lorenz flow.
5. Plot the strange attractor both in the original form (10.15) and in the proto-Lorenz form (11.17)



for the Lorenz parameter values $\sigma = 10$, $b = 8/3$, $\rho = 28$. Topologically, does it resemble more the Lorenz, or the Rössler attractor, or neither? (plot by J. Halcrow)

7. Show that a periodic orbit of the proto-Lorenz is either a periodic orbit or a relative periodic orbit of the Lorenz flow.
8. Show that if a periodic orbit of the proto-Lorenz is also periodic orbit of the Lorenz flow, their Floquet multipliers are the same. How do the Floquet multipliers of relative periodic orbits of the Lorenz flow relate to the Floquet multipliers of the proto-Lorenz?
- 9 What does the volume contraction formula (4.42) look like now? Interpret.
10. Show that the coordinate change (11.16) is the same as rewriting (11.15) in variables

$$(u, v) = (r^2 \cos 2\theta, r^2 \sin 2\theta),$$

i.e., squaring a complex number $z = x + iy$, $z^2 = u + iv$.

11. How is (11.17) related to the invariant polynomial basis of example 11.12 and exercise 11.15?

Chapter 12

Relativity for cyclists

Physicists like symmetry more than Nature

— Rich Kerswell



WHAT IF THE LAWS OF MOTION retain their form for a family of coordinate frames related by *continuous* symmetries? The finite groups intuition is of little use here.



example 12.1

p. 228

First of all, why worry about continuous symmetries? In physics, we usually assume isotropy, that is the laws of nature do not depend on where we are. In many cases (single or many body quantum mechanics, statistical physics, field theories etc.), the system studied is defined over an infinite or periodic domain. For linear problems of this kind, one takes care of the spatial dependence via a Fourier expansion (2.19), and solves the problem for each mode separately. In nonlinear theories one might start with a Fourier expansion, but the modes couple nonlinearly, and hence one needs to solve for all of them simultaneously. The two-modes system, which we shall introduce in sect. 12.4.3 and use for illustrations throughout this chapter and the next, is an example of a few modes truncation of such Fourier expansion. Figure 12.1 illustrates the effect continuous symmetry has on dynamics of this particular system. The strange attractor is a mess, and, as we shall demonstrate, what makes it messy is its continuous symmetry.

section 12.4.3

remark 12.2

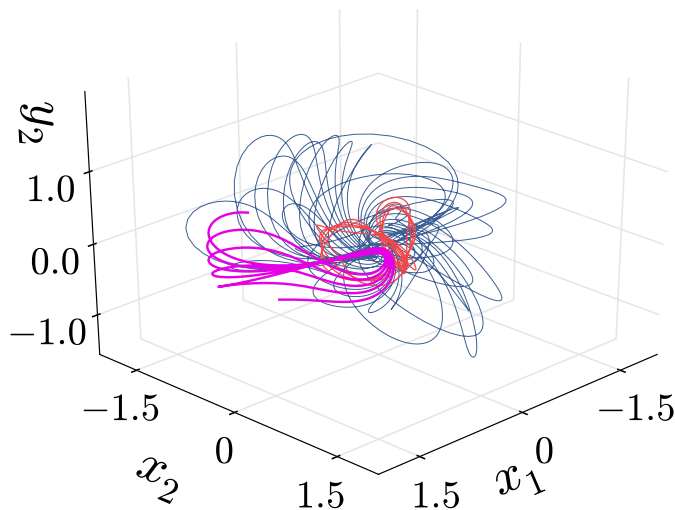


example 12.8

p. 230

We shall refer to the component of the dynamics along the continuous symmetry directions as a ‘drift’. In the presence of continuous symmetry, an orbit explores the manifold swept by combined action of the dynamics and the symmetry induced drifts. Further problems arise when we try to determine whether an orbit shadows another orbit (see figure 16.1 for a sketch of a close pass to a periodic orbit), or develop symbolic dynamics (partition the state space, as in

Figure 12.1: Several trajectories of the 4-dimensional two-modes system of example 12.8, a 3-dimensional projection: A long trajectory that originated close to the relative equilibrium TW_1 of the two-modes flow (12.40), with the starting point on its unstable manifold. The initial segment of this trajectory, which follows closely the orbit of TW_1 (see figure 12.5), is colored red; beyond that the trajectory falls onto the strange attractor (colored blue). Superimposed, in magenta, are four repeats of the shortest relative periodic orbit $\bar{1}$ (see figure 12.7 (b)). (N.B. Budanur)



chapter 14). Here a 1-dimensional trajectory is replaced by a $(N+1)$ -dimensional ‘sausage’, a dimension for each continuous symmetry (N being the total number of parameters specifying the continuous transformation, and ‘1’ for the time parameter t). How are we to measure distances between such objects? In this chapter and the next one, we shall learn how to develop visualizations of such flows, quotient symmetries, and offer computationally straightforward methods of reducing the dynamics to lower-dimensional, reduced state spaces. The methods should also be applicable to high-dimensional flows, such as translationally invariant fluid flows bounded by pipes or planes.



example 12.2
p. 228



example 12.3
p. 228



example 12.4
p. 228

Instead of writing yet [another tome on group theory](#), in what follows we continue to serve group theoretic nuggets on need-to-know basis, through a series of pedestrian examples (but take a slightly higher, cyclist road in the text proper).

12.1 Continuous symmetries

I’ve always hated the term ‘group orbit’
— John F. Zappatista

But first, a lightning review of the theory of Lie groups. The group-theoretical concepts of sect. 10.1 apply to compact continuous groups as well, and will not be repeated here.



example 12.5
p. 229



example 12.6
p. 229

Let G be a group, and $gM \rightarrow M$ a group action on the state space M . The $[d \times d]$ matrices g acting on vectors in the d -dimensional state space M form a

linear representation of the group G . If the action of every element g of a group G commutes with the flow

$$gv(x) = v(gx), \quad gf^\tau(x) = f^\tau(gx), \quad (12.1)$$

G is a symmetry of the dynamics, and, as in (10.4), the dynamics is said to be G -equivariant. 

In order to explore the implications of equivariance for the solutions of dynamical equations, we start by examining the way a compact Lie group acts on state space $\mathcal{M}$.

Definition: Group orbit For any $x \in \mathcal{M}$, the *group orbit* $\mathcal{M}_x$ of x is the set of all points that x is mapped to under the groups actions, 

$$\mathcal{M}_x = \text{Orb}(x) = \{gx \mid g \in G\}. \quad (12.2)$$

See page 184 and figure 12.2 (a).

Definition: Fixed-point subspace $\mathcal{M}_H$, or a ‘centralizer’ of a subgroup $H \subset G$, is the set of all state space points that are *H-fixed, point-wise* invariant under action of the subgroup

$$\mathcal{M}_H = \text{Fix}(H) = \{x \in \mathcal{M} \mid hx = x \text{ for all } h \in H\}. \quad (12.3)$$

Points in the *fixed-point subspace* $\mathcal{M}_G$ are fixed points of the full group action, i.e., points whose group orbit consists of only the point itself ($\mathcal{M}_x = \{x\}$). They are called *invariant points*,

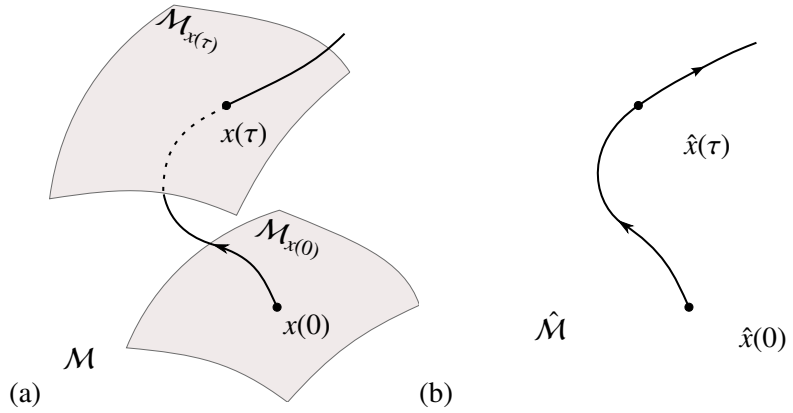
$$\mathcal{M}_G = \text{Fix}(G) = \{x \in \mathcal{M} \mid gx = x \text{ for all } g \in G\}. \quad (12.4)$$

If a point is an invariant point of the symmetry group by the definition of equivariance (12.1), the velocity at that point is also in $\mathcal{M}_G$. Thus, the trajectory through that point will remain in $\mathcal{M}_G$. $\mathcal{M}_G$ is disjoint from the rest of the state space since no trajectory can ever enter or leave it.

example 12.6

The time evolution itself is a noncompact 1-parameter Lie group, which means that time evolution and continuous symmetries can be considered on the same Lie group footing. For a given state space point x , a symmetry group of N continuous transformations together with the evolution in time sweeps out, in general, a smooth $(N+1)$ -dimensional manifold of equivalent solutions, see figure 12.3 (if the solution has a symmetry, the manifold may have a dimension less than $N+1$). For solutions for which the group orbit of x_p is periodic in time T_p , the group orbit sweeps out a *compact* invariant manifold $\mathcal{M}_p$. The simplest example is the $N = 0$, no symmetry case, where the invariant manifold $\mathcal{M}_p$ is the 1-torus traced out by a periodic trajectory p . If $\mathcal{M}$ is a smooth C^∞ manifold, and G is compact and acts smoothly on $\mathcal{M}$, the reduced state space can be realized as a ‘stratified manifold’, meaning that each group orbit (a ‘stratum’) is represented by a point in 

Figure 12.2: (a) The group orbit $\mathcal{M}_{x(0)}$ of state space point $x(0)$, and the group orbit $\mathcal{M}_{x(\tau)}$ reached by the trajectory $x(\tau)$ time τ later. As any point on the manifold $\mathcal{M}_{x(\tau)}$ is physically equivalent to any other, the state space is stratified into the union of group orbits. (b) Symmetry reduction $\mathcal{M} \rightarrow \hat{\mathcal{M}}$ replaces each full state space group orbit $\mathcal{M}_x$ by a single point $\hat{x} \in \hat{\mathcal{M}}$.



the reduced state space, see figure 12.2 and sect. 13.2. Generalizing the description of a non-wandering set of sect. 2.1.1, we say that for flows with continuous symmetries the non-wandering set Ω of dynamics (2.5) is the closure of the set of compact invariant manifolds $\mathcal{M}_p$. Without symmetries, we visualize the non-wandering set as a set of points; in the presence of a continuous symmetry, each such ‘point’ is a group orbit.

12.1.1 Lie groups for pedestrians

[...] which is an expression of consecration of angular momentum.

— Mason A. Porter’s student

Definition: A Lie group is a topological group G such that (i) G has the structure of a smooth differential manifold, and (ii) the composition map $G \times G \rightarrow G : (g, h) \rightarrow gh^{-1}$ is smooth, i.e., $\mathbb{C}^\infty$ differentiable.

Do not be mystified by this definition. Mathematicians also have to make a living. Historically, the theory of compact Lie groups that we will deploy here emerged as a generalization of the theory of $SO(2)$ rotations, i.e., Fourier analysis. A ‘smooth differential manifold’ includes objects like the circle of angles that parameterize continuous rotations in a plane, example 12.5, or the manifold swept by the three Euler angles that parameterize $SO(3)$ rotations.

An element of a Lie group continuously connected to identity can be written as

$$g(\phi) = e^{\phi \cdot \mathbf{T}}, \quad \phi \cdot \mathbf{T} = \sum_{a=1}^N \phi_a \mathbf{T}_a, \quad (12.5)$$

where $\phi \cdot \mathbf{T}$ is a *Lie algebra* element, and ϕ_a are the parameters of the transformation. Repeated indices are summed throughout this chapter, and the dot product refers to a sum over Lie algebra generators. We find it convenient to use bra-ket

notation for the Euclidean product of two real vectors $x, y \in \mathcal{M}$, i.e., indicate x -transpose times y by

$$\langle x|y \rangle = x^\top y = \sum_i^d x_i y_i. \quad (12.6)$$

Unitary transformations $\exp(\phi \cdot \mathbf{T})$ are generated by sequences of infinitesimal steps of form

$$g(\delta\phi) \simeq 1 + \delta\phi \cdot \mathbf{T}, \quad \delta\phi \in \mathbb{R}^N, \quad |\delta\phi| \ll 1, \quad (12.7)$$

where $\mathbf{T}_a$, the *generators* of infinitesimal transformations, are a set of linearly independent $[d \times d]$ anti-hermitian matrices, $(\mathbf{T}_a)^\dagger = -\mathbf{T}_a$, acting linearly on the d -dimensional state space $\mathcal{M}$. In order to streamline the exposition, we postpone discussion of combining continuous coordinate transformations with the discrete ones to sect. 12.2.1.

exercise 12.1

Unitary and orthogonal groups (as well as their subgroups) are defined as groups that preserve ‘length’ norms, $\langle gx|gx \rangle = \langle x|x \rangle$, and infinitesimally their generators (12.7) induce no change in the norm, $\langle \mathbf{T}_a x|x \rangle + \langle x|\mathbf{T}_a x \rangle = 0$, hence the Lie algebra generators $\mathbf{T}$ are antisymmetric for orthogonal groups, and antihermitian for unitary ones,

$$\mathbf{T}^\dagger = -\mathbf{T}. \quad (12.8)$$

For continuous groups, Lie algebra, i.e., the set of N generators $\mathbf{T}_a$ of infinitesimal transformations, takes the role that the $|G|$ group elements play in the theory of discrete groups. The flow field at the state space point x induced by the action of the group is given by the set of N *tangent fields* (see figure 12.3)

$$t_a(x)_i = (\mathbf{T}_a)_{ij} x_j, \quad (12.9)$$

which span the group *tangent space* at state space point x . The antisymmetry (12.8) of generators implies that the action of the group on vector x is locally normal to it,

$$\langle x|t_a(x) \rangle = 0. \quad (12.10)$$

A group tangent (12.9) is labelled by a pair of indices, as it is a vector both in the group tangent space and in the state space. We shall indicate by $\langle t_a(x)|t_b(y) \rangle$ the sum over state space inner product only, and by

$$\langle t(x)|t(y) \rangle = \sum_{a=1}^N \langle t_a(x)|t_a(y) \rangle = \langle x|\mathbf{T}^\dagger \cdot \mathbf{T} y \rangle \quad (12.11)$$

the sum over both group and spatial dimensions.

Any representation of a compact Lie group G is fully reducible, and invariant tensors constructed by contractions of $\mathbf{T}_a$ are useful for identifying irreps. The simplest such invariant is

$$\mathbf{T}^\top \cdot \mathbf{T} = \sum_{\alpha} C_2^{(\alpha)} \mathbf{1}^{(\alpha)}, \quad (12.12)$$

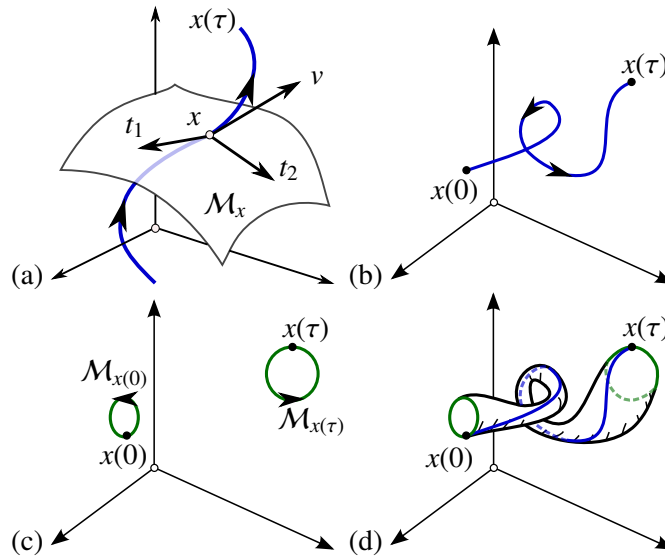


Figure 12.3: (a) In the presence of an N -continuous parameter symmetry, each state space point x owns $(N + 1)$ tangent vectors: one $v(x)$ along the time flow $x(\tau)$, and the N group tangents $t_1(x), t_2(x), \dots, t_N(x)$ along infinitesimal symmetry shifts, tangent to the N -dimensional group orbit $\mathcal{M}_x$. (b) Each point has a unique trajectory (blue) under time evolution. (c) Each point also belongs to a group orbit (green) of symmetry-related points. For $SO(2)$, this is topologically a circle. Any two points on a group orbit are physically equivalent, but may lie far from each other in state space. (d) Together, time-evolution and group actions trace out a “wurst” of physically equivalent solutions.

where $C_2^{(\alpha)}$ is a number called the quadratic Casimir for irrep labeled α , and $\mathbb{1}^{(\alpha)}$ is the identity on the α -irreducible subspace, 0 elsewhere. The dot product of two tangent fields is thus a sum weighted by Casimirs,

$$\langle t(x) | t(x') \rangle = \sum_{\alpha} C_2^{(\alpha)} x_i \delta_{ij}^{(\alpha)} x'_j. \quad (12.13)$$



example 12.7
p. 229



example 12.9
p. 231



The really interesting Lie groups are the non-abelian semisimple ones -but- as we will discuss nothing is more complicated than the abelian special orthogonal group $SO(2)$ of rotations in a plane, we shall not discuss the non-abelian case here.



fast track:
sect. 12.2, p. 218

Question 12.1. Alice wants to know

Q Why do you devote to Lie groups only two pages, while a book-length monograph would do it justice?

A ChaosBook tries its utmost to minimize the Gruppenpest jargon damage, which is a total turnoff to our intended audience of working plumbers and electricians. The sufferings of master Fabian Waleffe (see page 889) while reading chapter 10 *World in a mirror* are chicken feed in comparison to the continuous symmetry reduction nightmare that we will embark upon here.

section 12.1.1

appendix A1.6

chapter 13

12.1.2 Equivariance under infinitesimal transformations

A flow $\dot{x} = v(x)$ is G -equivariant (12.1), if symmetry transformations commute with time evolution

$$v(x) = g^{-1} v(g x), \quad \text{for all } g \in G. \quad (12.14)$$

For an infinitesimal transformation (12.7) the G -equivariance condition becomes 

$$v(x) = (1 - \phi \cdot \mathbf{T}) v(x + \phi \cdot \mathbf{T}x) + \dots = v(x) - \phi \cdot \mathbf{T}v(x) + \frac{dv}{dx} \phi \cdot \mathbf{T}x + \dots.$$

The $v(x)$ cancel, and ϕ_a are arbitrary. Denote the *group flow tangent field* at x by $t_a(x)_i = (\mathbf{T}_a)_{ij}x_j$. Thus the infinitesimal, Lie algebra G -equivariance condition is

$$t_a(v) - A(x) t_a(x) = 0, \quad (12.15)$$

where $A = \partial v / \partial x$ is the stability matrix (4.3). A learned remark: The directional derivative along direction ξ is $\lim_{t \rightarrow 0} (f(x + t\xi) - f(x))/t$. The left-hand side of (12.15) is the *Lie derivative* of the dynamical flow field v along the direction of the infinitesimal group-rotation induced flow $t_a(x) = \mathbf{T}_a x$,

$$\mathcal{L}_{t_a} v = \left(\mathbf{T}_a - \frac{\partial}{\partial y} (\mathbf{T}_a x) \right) v(y) \Big|_{y=x}. \quad (12.16)$$

The equivariance condition (12.15) states that the two flows, one induced by the dynamical vector field v , and the other by the group tangent field t , commute if their Lie derivatives (or the ‘Lie brackets’ or ‘Poisson brackets’) vanish.

exercise 12.9



example 12.10
p. 232



example 12.11
p. 232

Checking equivariance as a Lie algebra condition (12.15) is easier than checking it for global, finite angle rotations (12.14).

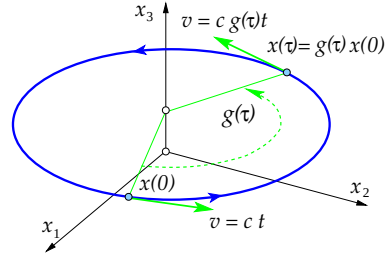
12.2 Symmetries of solutions

Let $v(x)$ be the dynamical flow, and f^τ the trajectory or ‘time- τ forward map’ of an initial point x_0 , 

$$\frac{dx}{dt} = v(x), \quad x(\tau) = f^\tau(x_0) = x_0 + \int_0^\tau d\tau' v(x(\tau')). \quad (12.17)$$

As discussed in sect. 11.1, solutions $x(\tau)$ of an equivariant system can satisfy all of the system’s symmetries, a subgroup of them, or have no symmetry at all. For a given solution $x(\tau)$, the subgroup that contains all symmetries that fix x (that satisfy $gx = x$) is called the isotropy (or stabilizer) subgroup of x . A generic ergodic trajectory $x(\tau)$ has no symmetry beyond the identity, so its isotropy group is $\{e\}$, but recurrent solutions often do. At the other extreme is equilibrium or steady solution (2.11), whose isotropy group is the full symmetry group G .

Figure 12.4: A *relative equilibrium orbit* starts out at some point $x(0)$, with the dynamical flow field $v(x) = c \cdot t(x)$ pointing along the group tangent space. For the $SO(2)$ symmetry depicted here, the flow traces out the group orbit of $x(0)$ in time $T = 2\pi/c$. An *equilibrium* lives either in the fixed $\text{Fix}(G)$ subspace (x_3 axis in this sketch), or on a group orbit as the one depicted here, but with zero angular velocity c . In that case the circle (in general, N -torus) depicts a continuous family of fixed equilibria, related only by the group action.



Definition: Equilibrium $x_{EQ} = \mathcal{M}_{EQ}$ is a fixed, time-invariant solution,

$$\begin{aligned} v(x_{EQ}) &= 0, \\ x(x_{EQ}, \tau) &= x_{EQ} + \int_0^\tau d\tau' v(x(\tau')) = x_{EQ}. \end{aligned} \quad (12.18)$$

An *equilibrium* with full symmetry,

$$g x_{EQ} = x_{EQ} \quad \text{for all } g \in G,$$

lies, by definition, in $\text{Fix}(G)$ subspace (12.3), for example the x_3 axis in figure 12.4(a). The multiplicity of such solution is one. An equilibrium x_{EQ} with symmetry G_{EQ} smaller than the full group G belongs to a group orbit G/G_{EQ} . If G is finite, there are $|G|/|G_{EQ}|$ equilibria in the group orbit. Otherwise, if G is continuous, then the group orbit of x is a continuous family of equilibria of dimension $\dim G - \dim G_{EQ}$. For example, if the angular velocity c in figure 12.4(b) equals zero, the group orbit consists of a circle of (dynamically static) equivalent equilibria.

Definition: Relative equilibrium solution $x_{TW}(\tau) \in \mathcal{M}_{TW}$: the dynamical flow field points along the group tangent field, with constant ‘angular’ velocity c , and the trajectory stays on the group orbit, see figure 12.4(a):

$$\begin{aligned} v(x) &= c \cdot t(x), \quad x \in \mathcal{M}_{TW} \\ x(\tau) &= g(-\tau c) x(0) = e^{-\tau c \cdot \mathbf{T}} x(0). \end{aligned} \quad (12.19)$$

A *traveling wave*

question 13.2

$$x(\tau) = g(-c\tau) x_{TW} = x_{TW} - c\tau, \quad c \in \mathbb{R}^d \quad (12.20)$$

is a special type of a relative equilibrium of equivariant evolution equations, where the action is given by translation (12.32), $g(y) x(0) = x(0) + y$. A *rotating wave* is another special case of relative equilibrium, where the action is angular rotation. By equivariance, all points on the group orbit are equivalent and the magnitude of the velocity c is the same everywhere along the orbit. Thus, a ‘traveling wave’ moves at a constant speed. For $N > 1$, the trajectory traces out a line within

Figure 12.5: $\{x_1, y_1, x_2\}$ plot of the two-modes system with initial point on the unstable manifold of TW_1 . In figure 12.1 this trajectory is integrated for a longer time, until it falls on to the strange attractor. (N.B. Budanur)

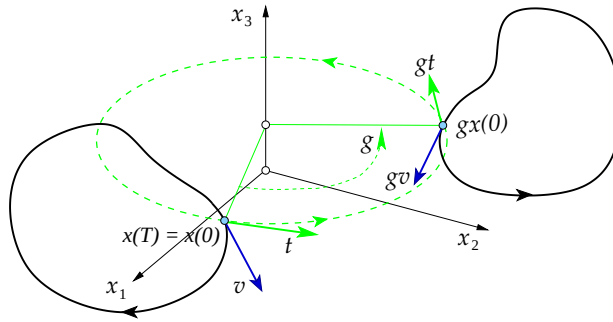
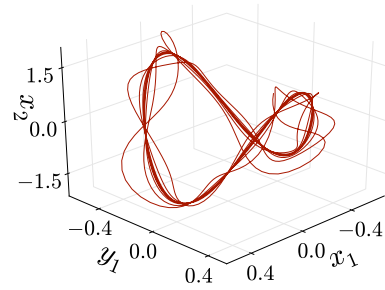


Figure 12.6: A periodic orbit starts out at $x(0)$ with the dynamical v and group tangent t flows pointing in different directions, and returns after time T to the initial point $x(0) = x(T)$. The group orbit of the temporal orbit of $x(0)$ sweeps out a $(1+N)$ -dimensional torus, a continuous family of equivalent periodic orbits, two of which are sketched here. For $SO(2)$ this is topologically a 2-torus.

the group orbit. As the c_a components are generically not in rational ratios, the trajectory explores the N -dimensional group orbit (12.2) quasi-periodically. In other words, the group orbit $g(\tau)x(0)$ coincides with the dynamical orbit $x(\tau) \in \mathcal{M}_{TW}$ and is also flow invariant.

Definition: Periodic orbit. Let x be a periodic point on the periodic orbit p of period T ,

$$f^T(x) = x, \quad x \in \mathcal{M}_p.$$

By equivariance, gx is another periodic point, with the orbits of x and gx either identical or disjoint.

If gx lands on the same orbit, g is an element of periodic orbit's symmetry group G_p . If the symmetry group is the full group G , we are back to (12.19), i.e., the periodic orbit is the group orbit traced out by a relative equilibrium. The other option is that the isotropy group is discrete, and the orbit segment $\{x, gx\}$ is pre-periodic (or eventually periodic), $x(0) = g_p x(T_p)$, where T_p is a fraction of the full period, $T_p = T/m$. Hence,

$$\begin{aligned} x(0) &= g_p x(T_p), & x &\in \mathcal{M}_p, & g_p &\in G_p \\ x(0) &= g_p^m x(m T_p) = x(T) = x(0). \end{aligned} \quad (12.21)$$

If the periodic solutions are disjoint, as in figure 12.6, their multiplicity (if G

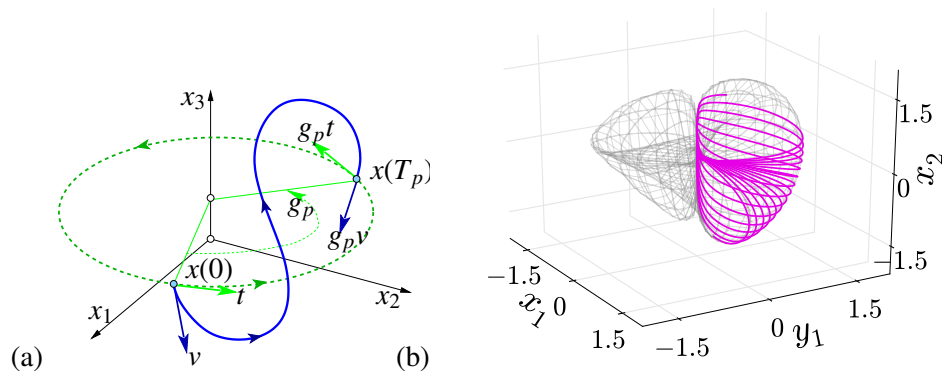


Figure 12.7: (a) A very idealized sketch: a relative periodic orbit starts out at $x(0)$ with the dynamical v and group tangent t flows pointing in different directions, and returns to the group orbit of $x(0)$ after time T_p at $x(T_p) = g_p x(0)$, a rotation of the initial point by g_p . For flows with continuous symmetry a generic relative periodic orbit (not pre-periodic to a periodic orbit) fills out ergodically what is topologically a torus, as in (b); if you are able to draw such a thing, kindly send us the figure. (b) The simplest example, 4-dimensional two-modes system of example 12.8, a 3-dimensional projection: 15 repeats of the shortest relative periodic orbit $\bar{I}$ (magenta) winding around the torus (here visualized as the gray wireframe). (N.B. Budanur)

is finite, see sect. 10.1), or the dimension of the manifold swept under the group action (if G is continuous) can be determined by applications of $g \in G$. They form a family of conjugate solutions (10.7),

$$\mathcal{M}_{g_p} = g \mathcal{M}_p g^{-1}. \quad (12.22)$$

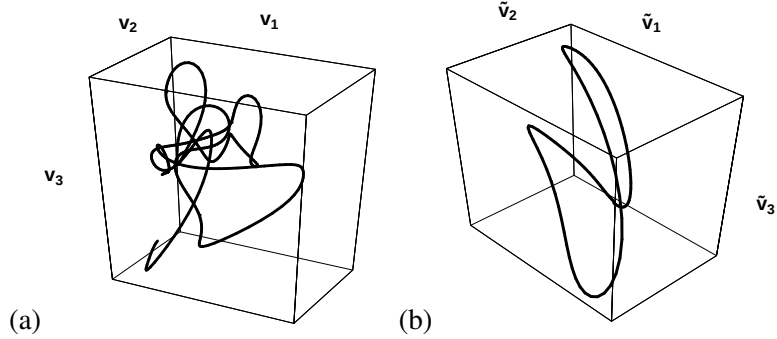
Definition: Relative periodic orbit p is an orbit $\mathcal{M}_p$ in state space $\mathcal{M}$ which exactly recurs

$$x_p(0) = g_p x_p(T_p), \quad x_p(\tau) \in \mathcal{M}_p, \quad (12.23)$$

at a fixed *relative period* T_p , but shifted by a fixed group action g_p which brings the endpoint $x_p(T_p)$ back into the initial point $x_p(0)$, see figure 12.7(a). The group action g_p parameters $\phi = (\phi_1, \phi_2, \dots, \phi_N)$ are referred to as ‘phases’, or ‘shifts’. In contrast to the pre-periodic (12.21), here the phases are irrational, and the trajectory sweeps out ergodically the group orbit without ever closing into a periodic orbit. For dynamical systems with only continuous (no discrete) symmetries, the parameters $\{t, \phi_1, \dots, \phi_N\}$ are real numbers, and the ratios π/ϕ_j are almost never rational. The likelihood of finding a periodic orbit for such system is zero, and such relative periodic orbits are almost never become periodic.

Relative periodic orbits are to periodic solutions what relative equilibria (traveling waves) are to equilibria (steady solutions). Equilibria satisfy $f^\tau(x) - x = 0$, and relative equilibria satisfy $f^\tau(x) - g(\tau)x = 0$ for any τ . In a co-moving frame, i.e., frame moving along the group orbit with velocity $v(x) = c \cdot t(x)$, the relative equilibrium appears as an equilibrium. Similarly, a relative periodic orbit is periodic in its mean velocity $c_p = \phi_p/T_p$ co-moving frame (see figure 12.8), but in the

Figure 12.8: A relative periodic orbit of Kuramoto-Sivashinsky flow projected on (a) the stationary state space coordinate frame $\{v_1, v_2, v_3\}$, traced for four periods T_p ; (b) the co-moving $\{\tilde{v}_1, \tilde{v}_2, \tilde{v}_3\}$ coordinate frame, moving with the mean angular velocity $c_p = \phi_p/T_p$ (from ref. [10]).



stationary frame its trajectory is quasiperiodic. A co-moving frame is helpful in visualizing a single ‘relative’ orbit, but useless for viewing collections of orbits, as each one drifts with its own angular velocity. We can only visualize all relative periodic orbits as periodic orbits by using global symmetry reductions, which is undertaken in sect. 13.2.

12.2.1 Discrete and continuous symmetries together

We expect to see relative periodic orbits because a trajectory that starts on and returns to a given torus of a symmetry equivalent solutions is unlikely to intersect it at the initial point, unless forced to do so by a discrete symmetry. This we will make explicit in sect. 13.2, where relative periodic orbits will be viewed as periodic orbits of reduced dynamics.

If, in addition to a continuous symmetry, one has a discrete symmetry which is not its subgroup, one does expect equilibria and periodic orbits. However, a relative periodic orbit can be pre-periodic if it is equivariant under a discrete symmetry. As in (12.21): If $g^m = 1$ is of finite order m , then the corresponding orbit is periodic with period mT_p . If g is not of a finite order, a relative periodic orbit is periodic only after a shift by g_p , as in (12.23). Morally, as it will be shown in chapter 25, such orbit is the true ‘prime’ orbit, i.e., the shortest segment that under action of G tiles the entire invariant submanifold $\mathcal{M}_p$.

Definition: Relative orbit $\mathcal{M}_{Gx}$ in state space $\mathcal{M}$ is the time evolved *group orbit* $\mathcal{M}_x$ of a state space point x . Hence, $\mathcal{M}_{Gx}$ is the set of all points that can be reached from x by all symmetry group actions and evolution of each in time.

$$\mathcal{M}_{x(\tau)} = \{gx(\tau) \mid t \in \mathbb{R}, g \in G\} . \quad (12.24)$$

In the presence of symmetry, an equilibrium is the set of all equilibria related by symmetries, and a relative periodic orbit is the hyper-surface traced by a trajectory in time T and all group actions, etc..

chapter 25

12.3 Stability



A spatial derivative of the equivariance condition (12.1) yields the matrix equivariance condition satisfied by the stability matrix (stated here both for the finite group actions, and for the infinitesimal, Lie algebra generators):

$$gA(x)g^{-1} = A(gx), \quad [\mathbf{T}_a, A] = \frac{\partial A}{\partial x} t_a(x). \quad (12.25)$$

For a flow within the fixed $\text{Fix}(G)$ subspace, $t(x)$ vanishes, and the symmetry imposes strong conditions on the perturbations out of the $\text{Fix}(G)$ subspace. In this subspace, stability matrix A commutes with Lie algebra generators $\mathbf{T}$, and the spectrum of its eigenvalues and eigenvectors is decomposed into irreps of the symmetry group. This we have already observed for EQ_0 of the Lorenz flow in example 11.8.

An infinitesimal symmetry group transformation maps the initial and end points of a finite trajectory into a set of nearby, slightly rotated equivalent points. We also expect the perturbations along the group orbit to be marginal, with unit eigenvalues. The argument is akin to (4.9), the proof of marginality of perturbations along a periodic orbit. Consider two nearby initial points separated by an N -dimensional infinitesimal group transformation (12.7): $\delta x_0 = g(\delta\phi)x_0 - x_0 = \delta\phi \cdot \mathbf{T}x_0 = \delta\phi \cdot t(x_0)$. By the commutativity of the group with the flow, $g(\delta\phi)f^\tau(x_0) = f^\tau(g(\delta\phi)x_0)$. When we expand both sides, keep the leading term in $\delta\phi$, and use the definition of the Jacobian matrix (4.5), $J^\tau(x_0)$ transports the N -dimensional group tangent space at $x(0)$ to the rotated tangent space at $x(\tau)$ at time τ :

$$t_a(\tau) = J^\tau(x_0) t_a(0), \quad t_a(\tau) = \mathbf{T}_a x(\tau). \quad (12.26)$$

For a relative periodic orbit, $g_p x(T_p) = x(0)$, at any point along cycle p , the group tangent vector $t_a(\tau)$ is an eigenvector of the Jacobian matrix with an eigenvalue of unit magnitude,

$$J_p t_a(x) = t_a(x), \quad J_p(x) = g_p J^{T_p}(x), \quad x \in \mathcal{M}_p. \quad (12.27)$$

For a relative equilibrium flow and group tangent vectors coincide, $v = c \cdot t(x)$. Dotting by the velocity c (i.e., summing over $c_a t_a$) the equivariance condition (12.15), $t_a(v) - A(x)t_a(x) = 0$, we get

$$(c \cdot \mathbf{T} - A)v = 0. \quad (12.28)$$

example 13.2

In other words, in the co-rotating frame, eigenvalues corresponding to group tangent are marginal, and velocity v is the corresponding right eigenvector.

Two successive points along the cycle separated by $\delta x_0 = \delta\phi \cdot t(\tau)$ have the same separation after a completed period $\delta x(T_p) = g_p \delta x_0$, hence eigenvalue of magnitude 1. In the presence of an N -dimensional Lie symmetry group, N eigenvalues equal unity.

Résumé

The message: If a dynamical system has symmetry, use it!

We conclude with a few general observations: Higher dimensional dynamics requires study of compact invariant sets of higher dimension than 0-dimensional equilibria and 1-dimensional periodic orbits studied so far. In sect. 2.1.1 we made an attempt to classify ‘all possible motions:’ (a) equilibria, (b) periodic orbits, (c) everything else. Now one can discern in the fog of dynamics an outline of a more serious classification - long time dynamics takes place on the closure of a set of all invariant compact sets preserved by the dynamics, and those are: (a) 0-dimensional equilibria M_{EQ} , (b) 1-dimensional periodic orbits M_p , (3) global symmetry induced N -dimensional relative equilibria M_{TW} , (c) $(N+1)$ -dimensional relative periodic orbits M_p , (d) terra incognita. We have some inklings for ‘terra incognita.’ For example, in symplectic symmetry settings, one finds KAM-tori; and, in general dynamical settings, one encounters *partially hyperbolic invariant M-tori*, isolated tori that are consequences of dynamics, not of global symmetry. These M -tori are harder to compute than anything we have attempted so far, as they cannot be represented by a single relative periodic orbit, but require a numerical computation of full M -dimensional compact invariant sets and their infinite-dimensional linearized Jacobian matrices, marginal in M dimensions, and hyperbolic in the rest. We expect partially hyperbolic invariant tori to play an important role in high-dimensional dynamics. In this chapter, we have focused on the simplest example of such compact invariant sets, where invariant tori are a robust consequence of a global continuous symmetry of dynamics. The direct product structure of a global symmetry that commutes with the flow enables us to reduce the dynamics to a desymmetrized $(d-1-N)$ -dimensional reduced state space M/G .

Relative equilibria and relative periodic orbits are the hallmark of systems with continuous symmetry. Amusingly, in this extension of ‘periodic orbit’ theory from unstable 1-dimensional closed periodic orbits to unstable $(N+1)$ -dimensional compact manifolds M_p invariant under continuous symmetries, there are either no or proportionally few periodic orbits. In the presence of a continuous symmetry, the likelihood of finding a periodic orbit is *zero*. Relative periodic orbits are almost never eventually periodic, i.e., they almost never lie on periodic trajectories in full state space, so looking for periodic orbits in systems with continuous symmetries is a fool’s errand.

However, dynamical systems are often equivariant under a combination of continuous symmetries and the discrete coordinate transformations of chapter 10, such as the orthogonal group $O(n)$. In the presence of discrete symmetries, relative periodic orbits within discrete symmetry-invariant subspaces are eventually periodic. Atypical as they are (no generic chaotic orbit can ever enter these discrete invariant subspaces), these orbits are important to periodic orbit theory, since the shortest orbits dominate, and their solutions tend to be the most symmetric.

chapter 25

Commentary

Remark 12.1. Ideal is not real. (continued from remark 10.1): The literature on symmetries in dynamical systems is immense, most of it deliriously unintelligible. Would it kill them [15, 16, 20, 24] to say ‘symmetry of orbit p ’ instead of carrying on about ‘isotropies, quotients, factors, normalizers, centralizers and stabilizers?’ Group action being ‘free, faithful, proper, regular?’ Symmetry-reduced state space being ‘orbifold?’ For the dynamical systems applications at hand we need only the basic Lie group facts, on the level of any standard group theory textbook [18]. As an introduction, we found the text written by Roger Penrose [27] to be both enjoyable and understandable. Bluman and Kumei [3] Chapter 2 offers a pedagogical introduction to Lie groups of transformations, and Nakahara [25] to Lie derivatives and brackets. The presentation given here is in part based on Siminos thesis [30] and ref. [31]. Term ‘stabilizer’ is used, for example, by Broer *et al.* [5] to refer to a periodic orbit with Z_2 symmetry; they say that the relative or pre-periodic segment is in this case called a ‘short periodic orbit.’ In Efstathiou [13] a subgroup of ‘short periodic orbit’ symmetries is referred to as a ‘nontrivial isotropy group or stabilizer.’ The reader is referred to the monographs of Golubitsky and Stewart [16], Hoyle [20], Olver [26], Bredon [4], and Krupa [22] for more depth and rigor than would be wise to wade into here.

Remark 12.2. Two-modes system. Dangelmayr [11], Armbruster, Guckenheimer and Holmes [1], Jones and Proctor [21], and Porter and Knobloch [28] (see Golubitsky *et al.* [17], Sect. XX.1) have investigated bifurcations in 1:2 resonance ODE normal form models to the third order in amplitude. Budanur *et al.* [7] studied (12.40) a particular case of Dangelmayr [11], and Porter and Knobloch [28] studied 2-Fourier mode $SO(2)$ -equivariant ODEs, as a relatively simple application of the periodic orbit theory to a system with continuous symmetry. In this chapter and the next, we will use this model, which we will refer to as ‘two-modes’ system, to illustrate the effects of continuous symmetry on the dynamics and symmetry reduction by the method of slices.

Remark 12.3. Modulated traveling waves. When a ‘traveling wave’ goes unstable through a Hopf bifurcation, the resulting motion resembles the initial traveling wave weakly periodically ‘modulated’ in time, hence such relative periodic orbit is often called a *modulated traveling wave* (MTW), or a travelling, beating wave [29]. These were studied, for instance, by Armbruster *et al.* [1], and a detailed computation of numerous bifurcation branches of these solutions was presented by Brown and Kevrekidis [6]. They find quasiperiodic secondary Hopf bifurcations. In chaos unstable recurrent motions typically arise from stretching and folding mechanisms. So for our purposes, ‘MTW’ is too narrow a concept, merely a particular case of a relative periodic orbit.

References

- [1] D. Armbruster, J. Guckenheimer, and P. Holmes, “Heteroclinic cycles and modulated travelling waves in systems with $O(2)$ symmetry”, *Physica D* **29**, 257–282 (1988).
- [2] N. L. Balász and A. Voros, “Chaos on the pseudosphere”, *Phys. Rep.* **143**, 109–240 (1986).

- [3] G. W. Bluman and S. Kumei, *Symmetries and Differential Equations* (Springer, New York, 1989).
- [4] G. Bredon, *Introduction to Compact Transformation Groups* (Academic, New York, 1972).
- [5] H. W. Broer, I. Hoveijn, G. A. Lunter, and G. Vegter, *Bifurcations in Hamiltonian Systems: Computing Singularities by Gröbner Bases* (Springer, New York, 2003).
- [6] H. S. Brown and I. G. Kevrekidis, “Modulated traveling waves for the Kuramoto-Sivashinsky equation”, in *Pattern Formation: Symmetry Methods and Applications*, edited by J. Chadam, M. Golubitsky, W. Langford, and B. Wetton (AMS, Providence, RI, 1996), pp. 45–66.
- [7] N. B. Budanur, D. Borrero-Echeverry, and P. Cvitanović, “Periodic orbit analysis of a system with continuous symmetry - A tutorial”, *Chaos* **25**, 073112 (2015).
- [8] G. J. Chandler and R. R. Kerswell, “Invariant recurrent solutions embedded in a turbulent two-dimensional Kolmogorov flow”, *J. Fluid Mech.* **722**, 554–595 (2013).
- [9] F. Christiansen, P. Cvitanović, and V. Putkaradze, “Hopf’s last hope: Spatiotemporal chaos in terms of unstable recurrent patterns”, *Nonlinearity* **10**, 55–70 (1997).
- [10] P. Cvitanović, R. L. Davidchack, and E. Siminos, “On the state space geometry of the Kuramoto-Sivashinsky flow in a periodic domain”, *SIAM J. Appl. Dyn. Syst.* **9**, 1–33 (2010).
- [11] G. Dangelmayr, “Steady-state mode interactions in the presence of $O(2)$ -symmetry”, *Dyn. Sys.* **1**, 159–185 (1986).
- [12] T. Dombre, U. Frisch, J. M. Greene, M. Hénon, A. Mehr, and A. M. Soward, “Chaotic streamlines in the ABC flows”, *J. Fluid Mech.* **167**, 353–391 (1986).
- [13] K. Efsthathiou, *Metamorphoses of Hamiltonian systems with Symmetries* (Springer, New York, 2005).
- [14] J. F. Gibson, J. Halcrow, and P. Cvitanović, “Equilibrium and traveling-wave solutions of plane Couette flow”, *J. Fluid Mech.* **638**, 243–266 (2009).
- [15] R. Gilmore and C. Letellier, *The Symmetry of Chaos* (Oxford Univ. Press, Oxford, 2007).
- [16] M. Golubitsky and I. Stewart, *The Symmetry Perspective* (Birkhäuser, Boston, 2002).
- [17] M. Golubitsky, I. Stewart, and D. G. Schaeffer, *Singularities and Groups in Bifurcation Theory*, Vol. 2 (Springer, New York, 1988).
- [18] M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Dover, New York, 1962).
- [19] A. Hönig and D. Wintgen, “Spectral properties of strongly perturbed Coulomb systems: Fluctuation properties”, *Phys. Rev. A* **39**, 5642–5657 (1989).

- [20] R. Hoyle, *Pattern Formation: An Introduction to Methods* (Cambridge Univ. Press, Cambridge UK, 2006).
- [21] C. A. Jones and M. R. E. Proctor, “Strong spatial resonance and travelling waves in Benard convection”, *Phys. Lett. A* **121**, 224–228 (1987).
- [22] M. Krupa, “Bifurcations of relative equilibria”, *SIAM J. Math. Anal.* **21**, 1453–1486 (1990).
- [23] B. Lauritzen, “Discrete symmetries and the periodic-orbit expansions”, *Phys. Rev. A* **43**, 603–606 (1991).
- [24] J. E. Marsden and T. S. Ratiu, *Introduction to Mechanics and Symmetry* (Springer, New York, 1999).
- [25] M. Nakahara, *Geometry, Topology and Physics* (Inst. of Physics Publ., Bristol, 1990).
- [26] P. J. Olver, *Classical Invariant Theory* (Cambridge Univ. Press, Cambridge UK, 1999).
- [27] R. Penrose, *The Road to Reality - A Complete Guide to the Laws of the Universe* (A. A. Knopf, New York, 2005).
- [28] J. Porter and E. Knobloch, “Dynamics in the 1:2 spatial resonance with broken reflection symmetry”, *Physica D* **201**, 318–344 (2005).
- [29] C. W. Rowley, I. G. Kevrekidis, J. E. Marsden, and K. Lust, “Reduction and reconstruction for self-similar dynamical systems”, *Nonlinearity* **16**, 1257–1275 (2003).
- [30] E. Siminos, *Recurrent Spatio-temporal Structures in Presence of Continuous Symmetries*, PhD thesis (School of Physics, Georgia Inst. of Technology, Atlanta, 2009).
- [31] E. Siminos and P. Cvitanović, “Continuous symmetry reduction and return maps for high-dimensional flows”, *Physica D* **240**, 187–198 (2011).
- [32] A. P. Willis, P. Cvitanović, and M. Avila, “Revealing the state space of turbulent pipe flow by symmetry reduction”, *J. Fluid Mech.* **721**, 514–540 (2013).

12.4 Examples

Example 12.1. Invariance under fractional rotations. Consider a velocity field $v(x)$ equivariant (12.1) under discrete cyclic subgroup $C_m = \{e, d, d^2, \dots, d^{m-1}\}$ of $SO(2)$ rotations by $2\pi/m$,

$$dv(x) = v(dx), \quad (d)^m = e.$$

[exercise 12.2](#)

The field $v(x)$ on the fundamental domain $2\pi/m$ is now a tile, whose m copies tile the entire domain. It is periodic on the fundamental domain and, thus, has Fourier expansion with Fourier modes $\cos(2\pi m j x)$, $\sin(2\pi m j x)$. The Fourier expansion on the full interval $(0, 2\pi)$ cannot have any other modes, as they would violate the C_m symmetry. This means that $SO(2)$ always has an infinity of discrete subgroups $C_2, C_3, \dots, C_m, \dots$; for each subgroup, the non-vanishing coefficients are only Fourier modes whose wave numbers are multiples of m .

[return to p. 212](#)

12.4.1 Symmetries of iconic fluid flows

Example 12.2. Continuous symmetries of the plane Couette flow. (continued from example 10.10) Every solution of Navier-Stokes equations belongs, by the $SO(2) \times O(2)$ symmetry, to a 2-torus T^2 of equivalent solutions. Furthermore, these tori are interrelated by a discrete D_2 group of spanwise and streamwise flips of the flow cell. (continued in example 12.3)

[return to p. 213](#)

Example 12.3. Relative orbits in the plane Couette flow. (continued from example 12.2) Translational symmetry allows for relative equilibria (traveling waves), characterized by a fixed profile Eulerian velocity $u_{TW}(x)$ moving with constant velocity c , i.e.

$$u(x, \tau) = u_{TW}(x - c\tau). \quad (12.29)$$

As the plane Couette flow is bound by two counter-moving planes, it is easy to see where the relative equilibrium (traveling wave) solutions come from. A relative equilibrium solution hugs close to one of the walls and drifts along with constant velocity, while moving slower than the wall and maintaining its shape. A relative periodic solution is a solution that recurs at time T_p with exactly the same disposition of the Eulerian velocity fields over the entire cell, but shifted by a 2-dimensional (streamwise, spanwise) translation g_p . By discrete symmetries these solutions come in counter-traveling pairs $u_q(x - c\tau)$, $-u_q(-x + c\tau)$. For example, for each one drifting along with the upper wall, there is a counter-moving one drifting along with the lower wall. Discrete symmetries also imply existence of strictly stationary solutions, or ‘standing waves’. For example, a solution with velocity fields antisymmetric under reflection through the midplane has equal flow velocities in opposite directions, and is thus an equilibrium stationary in time.

[return to p. 213](#)

Example 12.4. Traveling, rotating waves. Names ‘traveling waves’, and ‘rotating waves’ are descriptive of solutions of some PDEs with simple continuous symmetries.

For example, complex Ginzburg Landau equation is equivariant under the action of the group $g(\theta, y) \in G = S^1 \times \mathbb{R}$ on $u(x) \in \mathbb{R}^2$, given by translation in the domain and the rotation of $u(x)$,

$$g(\theta, y)u(x) = R(\theta)u(x + y), \quad R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \quad (12.30)$$

Hence complex Ginzburg Landau equation allows for *rotating wave* solutions of form $u(x, t) = R(-\omega t)\hat{u}(x - ct)$ with fixed profile $\hat{u}(x)$, velocity c and angular velocity ω . Traveling waves are typical of translationally invariant systems such as the plane Couette flow, example 12.3.

[return to p. 213](#)

12.4.2 Special orthogonal group SO(2)

Example 12.5. Special orthogonal group SO(2). (or S^1) is a group of length-preserving rotations in a plane. ‘Special’ refers to the requirement that $\det g = 1$, in contradistinction to the orthogonal group $O(n)$, which allows for length-preserving inversions through the origin with $\det g = -1$. A group element can be parameterized by angle ϕ , with the group multiplication law $g(\phi')g(\phi) = g(\phi' + \phi)$, and its action on smooth periodic functions $u(\phi + 2\pi) = u(\phi)$ generated by

$$g(\phi') = e^{\phi' \mathbf{T}}, \quad \mathbf{T} = \frac{d}{d\phi}. \quad (12.31)$$

Expand the exponential, apply it to a differentiable function $u(\phi)$, and you will recognize a Taylor series. So $g(\phi')$ shifts the coordinate by ϕ' , $g(\phi')u(\phi) = u(\phi' + \phi)$. (Continued in example 13.1)

[return to p. 213](#)

Example 12.6. Translation group. Differential operator $\mathbf{T}$ in (12.31) is reminiscent of the generator of spatial translations. The ‘constant velocity field’ $v(x) = v = c \cdot \mathbf{T}$ acts on x_j by replacing it by the velocity vector c_j . It is easy to verify by Taylor expanding a function $u(x)$ that the time evolution is nothing but a coordinate translation by (time $\times$ velocity):

$$e^{-\tau c \cdot \mathbf{T}} u(x) = e^{-\tau c \cdot \frac{d}{dx}} u(x) = u(x - \tau c). \quad (12.32)$$

As x is a point in the Euclidean $\mathbb{R}^d$ space, the group is not compact. A sequence of time steps in time evolution always forms an abelian Lie group, *albeit* never as trivial as this free ballistic motion.

If the group actions consist of N rotations which commute, for example act on an N -dimensional cell with periodic boundary conditions, the group is an abelian group that acts on a torus T^N .

[return to p. 213](#)

Example 12.7. SO(2) irreps. (Continued from example 12.5) Consider the action (12.31) of the one-parameter rotation group SO(2) on a smooth periodic function $u(\phi + 2\pi) = u(\phi)$ defined on a 1D-dimensional configuration space domain $x \in [0, 2\pi)$. The state space matrix representation of the SO(2) counter-clockwise (right-handed) rotation $g(\phi')u(\phi) = u(\phi + \phi')$ by angle ϕ' is block-diagonal, acting on the k th Fourier coefficient pair (x_k, y_k) in the Fourier series (2.19),

$$u(\phi) = x_0 + \sum_{k=1}^{\infty} (x_k \cos k\phi + y_k \sin k\phi). \quad (12.33)$$

by multiplication by

$$g^{(k)}(\phi') = \begin{pmatrix} \cos k\phi' & -\sin k\phi' \\ \sin k\phi' & \cos k\phi' \end{pmatrix}, \quad \mathbf{T}^{(k)} = \begin{pmatrix} 0 & -k \\ k & 0 \end{pmatrix}, \quad (12.34)$$

where $\mathbf{T}^{(k)}$ is the k th Fourier mode Lie group generator. The SO(2) group tangent (12.9) to state space point $u(\phi)$ on the k th invariant subspace is

$$t^{(k)}(u) = k \begin{pmatrix} -y_k \\ x_k \end{pmatrix}. \quad (12.35)$$

The L^2 norm of $t(u)$ is weighted by the SO(2) quadratic Casimir (12.12), $C_2^{(k)} = k^2$,

$$\langle t(u)^\top | t(u) \rangle = \oint \frac{d\phi}{2\pi} u(\phi)^\top \mathbf{T}^\top \mathbf{T} u(2\pi - \phi) = \sum_{k=1}^{\infty} k^2 (x_k^2 + y_k^2), \quad (12.36)$$

and converges only for sufficiently smooth $u(\phi)$. What does that mean? We saw in (12.32) that $\mathbf{T}$ generates translations, and by (12.34) the velocity of the k th Fourier mode is k times higher than for the $k = 1$ component. If $|u^{(k)}|$ does not fall off faster than $1/k$, the action of SO(2) is overwhelmed by the high Fourier modes.

[return to p. 217](#)

12.4.3 Two-modes SO(2)-equivariant flow

Example 12.8. Two-modes flow. Consider the pair of U(1)-equivariant complex ODEs

$$\begin{aligned} \dot{z}_1 &= (\mu_1 - i e_1) z_1 + a_1 z_1 |z_1|^2 + b_1 z_1 |z_2|^2 + c_1 \bar{z}_1 z_2 \\ \dot{z}_2 &= (\mu_2 - i e_2) z_2 + a_2 z_2 |z_1|^2 + b_2 z_2 |z_2|^2 + c_2 z_1^2, \end{aligned} \quad (12.37)$$

with z_1, z_2 complex, and all parameters real valued.

The two-modes system, which we use for illustrations throughout this chapter and the next, is an example of a few-modes truncation of a Fourier expansion, truncated in such a way that the model exhibits the same symmetry structure as many nonlinear field problems, while being drastically simpler to study.

We shall refer to this toy model as the two-modes system. It belongs to the family of simplest ODE systems that we know. This two-modes system has the following characteristics: (a) behaves in a continuous U(1) / SO(2) manner with no discrete symmetry (if at least one of $e_j \neq 0$), (b) models ‘weather’, in the same sense that Lorenz equation models ‘weather’, (c) exhibits chaotic dynamics, (d) can be easily visualized, in the dimensionally lowest possible setting required for chaotic dynamics, with the full state space of dimension $d = 4$, and the SO(2)-reduced dynamics taking place in 3 dimensions, and (e) reduces symmetry by a single global slice hyperplane through the method of slices technique.

For parameters far from the bifurcation values, this is merely a toy model with no physical interpretation, just like the iconic Lorenz flow (2.25). We use it to illustrate the effects of continuous symmetry on chaotic dynamics. We have not found a second

order truncation of such models that exhibits interesting dynamics. Hence, this model is third order in amplitude and has an unreasonably high number of parameters. After some experimentation we fix or set to zero various parameters, and in the numerical examples that follow, we settle for parameters set to

$$\begin{aligned}\mu_1 &= -2.8, \mu_2 = 1, e_1 = 0, e_2 = 1, \\ a_1 &= -1, a_2 = -2.66, b_1 = 0, b_2 = 0, c_1 = -7.75, c_2 = 1,\end{aligned}\quad (12.38)$$

unless explicitly stated otherwise. For these parameter values, the system exhibits chaotic behavior. Experiment. If you find a more interesting behavior for some other parameter values, please let us know. The simplified system of equations can now be written as a 3-parameter $\{\mu_1, c_1, a_2\}$ two-modes system,

$$\begin{aligned}\dot{z}_1 &= \mu_1 z_1 - z_1 |z_1|^2 + c_1 \bar{z}_1 z_2 \\ \dot{z}_2 &= (1 - i) z_2 + a_2 z_2 |z_1|^2 + z_1^2.\end{aligned}\quad (12.39)$$

In order to numerically integrate and visualize the flow, we recast the equations in real variables by substitution $z_1 = x_1 + i y_1, z_2 = x_2 + i y_2$. The two-modes system (12.37) is now a set of four coupled ODEs

$$\begin{aligned}\dot{x}_1 &= (\mu_1 - r^2) x_1 + c_1 (x_1 x_2 + y_1 y_2), & r^2 &= x_1^2 + y_1^2 \\ \dot{y}_1 &= (\mu_1 - r^2) y_1 + c_1 (x_1 y_2 - x_2 y_1) \\ \dot{x}_2 &= x_2 + y_2 + x_1^2 - y_1^2 + a_2 x_2 r^2 \\ \dot{y}_2 &= -x_2 + y_2 + 2 x_1 y_1 + a_2 y_2 r^2.\end{aligned}\quad (12.40)$$

Try integrating (12.40) with random initial conditions and long times, times far beyond when initial transients have died out. What is wrong with this picture? It is a mess. As we shall show here, the attractor is built up by a nice ‘stretch & fold’ action, but that is totally hidden from the view by the continuous symmetry induced drifts. In the rest of the examples for this chapter and the next, we shall investigate various ways of ‘quotienting’ this $SO(2)$ symmetry and reducing dynamics to a 3-dimensional symmetry state space. We shall not rest until we attain the simplicity and bliss of a 1-dimensional return map. (N.B. Budanur and P. Cvitanović)

[return to p. 212](#)

Example 12.9. $SO(2)$ rotations for two-modes system. Substituting the Lie algebra generator

$$\mathbf{T} = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 \\ 0 & 0 & 2 & 0 \end{pmatrix}\quad (12.41)$$

acting on a 4-dimensional state space (12.40) into (12.5) yields a finite angle $SO(2)$ rotation:

$$g(\phi) = \begin{pmatrix} \cos \phi & -\sin \phi & 0 & 0 \\ \sin \phi & \cos \phi & 0 & 0 \\ 0 & 0 & \cos 2\phi & -\sin 2\phi \\ 0 & 0 & \sin 2\phi & \cos 2\phi \end{pmatrix}.\quad (12.42)$$

Itinerary	$(x_{p,1}, y_{p,1}, x_{p,2}, y_{p,2})$	Period
1	(0.4525719, 0.0, 0.0509257, 0.0335428)	3.6415120
01	(0.4517771, 0.0, 0.0202026, 0.0405222)	7.3459412
0111	(0.4514665, 0.0, 0.0108291, 0.0424373)	14.6795175
01101	(0.4503967, 0.0, -0.0170958, 0.0476009)	18.3874094

Table 12.1: Several short relative periodic orbits of the two-modes system: itineraries, a periodic point in a Poincaré section for each orbit, the period.

From (12.34) we see that the action of $SO(2)$ on the complex Lorenz equations state space decomposes into $m = 0$ G -invariant $m = 1$ and $m = 2$ subspaces.

The generator $\mathbf{T}$ is indeed anti-hermitian, $\mathbf{T}^\dagger = -\mathbf{T}$, and the group is compact; its elements are parameterized by $\phi \bmod 2\pi$. Locally, at $x \in \mathcal{M}$, the infinitesimal action of the group is given by the group tangent field $t(x) = \mathbf{T}x = (-y_1, x_1, -2y_2, 2x_2)$. In other words, the flow induced by the group action is normal to the radial direction in the (x_1, y_1) and (x_2, y_2) planes.

[return to p. 217](#)

Example 12.10. Equivariance of the two-modes system. That two-modes (12.40) is equivariant under $SO(2)$ rotations (12.42) can be checked by substituting the Lie algebra generator (12.41) and the stability matrix (4.3) for two-modes (12.40), $A =$

$$\begin{pmatrix} \mu_1 - 3x_1^2 + c_1x_2 - y_1^2 & c_1y_2 - 2x_1y_1 & c_1x_1 & c_1y_1 \\ c_1y_2 - 2x_1y_1 & \mu_1 - x_1^2 - c_1x_2 - 3y_1^2 & -c_1y_1 & c_1x_1 \\ 2x_1 + 2a_2x_1x_2 & 2a_2x_2y_1 - 2y_1 & 1 + a_2(x_1^2 + y_1^2) & 1 \\ 2y_1 + 2a_2x_1y_2 & 2x_1 + 2a_2y_1y_2 & -1 & 1 + a_2(x_1^2 + y_1^2) \end{pmatrix}. \quad (12.43)$$

into the equivariance condition (12.15). Considering that $t(v)$ depends on the full set of equations (12.40), and $A(x)$ is only its linearization, this is not an entirely trivial statement. (N.B. Budanur)

[return to p. 218](#)

Example 12.11. How contracting is the two-modes flow? For the parameter values (12.39) the flow is strongly volume contracting (4.29),

$$\partial_i v_i = \sum_{i=1}^4 \lambda_i(x, t) = \text{tr } A(x) = 2[1 + \mu_1 - (2 - a_2)r^2] = -3.6 - 9.32r^2. \quad (12.44)$$

Note that this quantity depends on the full state space coordinates only through the $SO(2)$ -invariant r^2 , so the volume contraction rate is symmetry-invariant characterization of the flow, as it should be. The shortest relative periodic orbit $\bar{1}$ has period $T_1 = 3.64 \dots$ and typical $r^2 \approx 1$, (see table 12.1), so in one period a neighborhood of the relative periodic orbit is contracted by factor $\approx \exp(T_1 \text{tr } A(x)) \approx 3.7 \times 10^{-21}$. This is an insanely contracting flow; if we start with mm^4 cube around a periodic point, this volume (remember, two directions are marginal) shrinks to $\approx mm \times mm \times 10^{-11} mm \times 10^{-11} mm \approx mm \times mm \times \text{fermi} \times \text{fermi}$. Diameter of a proton is a couple of fermis. This strange attractor is *thin*!

[return to p. 218](#)

12.5 Exercises

- 12.1. **SO(2) rotations in a plane:** Show by exponentiation (12.5) that the SO(2) Lie algebra element $\mathbf{T}$ generates rotation g in a plane,

$$\begin{aligned} g(\theta) &= e^{\mathbf{T}\theta} = \cos \theta \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sin \theta \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}. \end{aligned} \quad (12.45)$$

- 12.2. **Invariance under fractional rotations.** Argue that if the isotropy group of the velocity field $v(x)$ is the discrete subgroup C_m of SO(2) rotations about an axis (let's say the 'z-axis'),

$$dv(x) = v(dx) = v(x), \quad (d)^m = e,$$

the only non-zero components of Fourier-transformed equations of motion are a_{jm} for $j = 1, 2, \dots$. Argue that the Fourier representation is then the quotient map of the dynamics, $\mathcal{M}/C_m$. (Hint: this sounds much fancier than what is - think first of how it applies to the Lorenz system and the 3-disk pinball.)

- 12.3. **U(1) equivariance of two-modes system for finite angles:** Show that the vector field in two-modes (12.37) is equivariant under (12.5), the unitary group U(1) acting on $\mathbb{R}^4 \cong \mathbb{C}^2$ by

$$g(\theta)(z_1, z_2) = (e^{i\theta} z_1, e^{i2\theta} z_2), \quad \theta \in [0, 2\pi). \quad (12.46)$$

- 12.4. **SO(2) equivariance of two-modes system for finite angles:** Show that two-modes (12.40) are equivariant under rotation for finite angles.

- 12.5. **Stability matrix of two-modes system:** Compute the stability matrix (12.43) for two-modes system (12.40).

- 12.6. **SO(2) equivariance of two-modes system for infinitesimal angles.** Show that two-modes equations are equivariant under infinitesimal SO(2) rotations. Compute the volume contraction rate (4.29), verify (12.44). Period of the shortest relative periodic orbit of this system is $T_1 = 3.6415120$. How much does a small volume centered on the relative periodic orbit contract in that time?

- 12.7. **Integrate the two-modes system:** Integrate (12.40) and plot a long trajectory of two-modes in the 4d state space, (x_1, y_1, y_2) projection, as in Figure 12.1.

- 12.8. **Classify possible symmetries of solutions for your research problem.** Classify types of solutions you expect in your research problem by their symmetries. Literature examples: plane Couette flow [14], pipe flow (sect. 2.2 and appendix A of ref. [32]), Kuramoto-Sivashinsky (see symmetry discussions of refs. [9, 10], and probably many better papers out there that we are less familiar with), Euclidean symmetries of doubly-periodic 2D models of cardiac tissue, 2D Kolmogorov flow [8], two-modes flow (Dangelmayr [11]; Armbruster, Guckenheimer and Holmes [1]; Jones and Proctor [21]; Porter and Knobloch [28]; Golubitsky *et al.* [17], Sect. XX.1), 2D ABC flow [12]; perturbed Coulomb systems [19]; systems with discrete symmetries [2, 23]; example 10.5 reflection symmetric 1d map; example 8.6 Hamiltonian Hénon map; Hamiltonian Lozi map, etc..

12.9. Discover the equivariance of a given flow:

Suppose you were given two-modes system, but nobody told you that the equations are $SO(2)$ -equivariant. More generally, you might encounter a flow without realizing that it has a continuous symmetry - how would you discover it?

Chapter 13

Slice & dice

Physicists like symmetry more than Nature
— Rich Kerswell



IF THE SYMMETRY IS CONTINUOUS, the notion of ‘fundamental domain’ is not applicable. Instead, the dynamical system should be reduced to a lower-dimensional, desymmetrized system, with ‘ignorable’ coordinates separated out (but not forgotten).

We shall describe here two ways of reducing a continuous symmetry. In the ‘method of slices’ of sect. 13.2, we slice the state space in such a way that an entire class of symmetry-equivalent points is represented by a single point. In the Hilbert polynomial basis approach of sect. 13.6, we replace the equivariant dynamics by the dynamics rewritten in terms of invariant coordinates. In either approach we retain the option of ‘post-processing’, i.e., computing in the original coordinates, and then, when done, projecting the solution onto the symmetry reduced state space.

In the method of slices, symmetry reduction is achieved by cutting the group orbits with a finite set of slice hyperplanes, one for each continuous group parameter, with each group orbit of symmetry-equivalent points represented by a single point, its intersection with the slice. The procedure is akin to (but distinct from) cutting across continuous-time parameterized trajectories by means of Poincaré sections. As is the case for Poincaré sections, choosing a ‘good’ slice is a dark art. We describe two strategies: (i) Foliation of state space by group orbits is a purely group-theoretic phenomenon that has nothing to do with dynamics, so we construct slices based on a decomposition of state space into irreducible linear representations of the symmetry group G , and (ii) Nonlinear dynamics strongly couples such linear symmetry eigenmodes, so locally optimal slices should be constructed from physically important recurrent states, or ‘templates’. Our guiding principle is to choose a slice such that the distance between a ‘template’ state $\hat{x}'$ and nearby group orbits is *minimized*; i.e., we identify the point $\hat{x}$ on the group orbit (12.2) of a nearby state x , which is the closest match to the template point $\hat{x}'$.



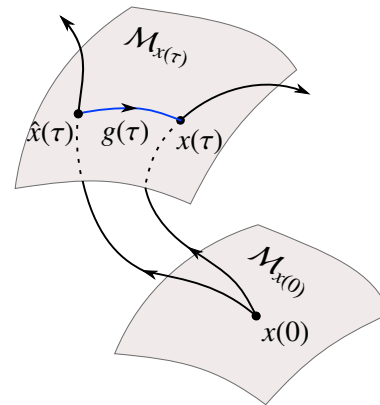


Figure 13.1: The freedom to pick a moving frame: A point x on the full state space trajectory $x(\tau)$ is equivalent up to a group rotation $g(\tau)$ to the point $\hat{x}$ on the curve $\hat{x}(\tau)$, if the two points belong to the same group orbit $\mathcal{M}_{x(\tau)}$, see (12.2).

We start our discussion by explaining the freedom of redefining dynamics in a moving frame.

13.1 Moving frames

Even a dead fish can go with the flow.
— Jim Hightower, Texas politician

The idea: As the symmetries commute with dynamics, we can evolve a solution $x(\tau)$ for as long as we like and then rotate it to any equivalent point (see figure 13.1) on its group orbit. We can map each point along any solution $x(\tau)$ to the unique representative $\hat{x}(\tau)$ of the associated group orbit equivalence class, by a coordinate transformation



$$x(\tau) = g(\tau) \hat{x}(\tau). \quad (13.1)$$

Equivariance guarantees that the two states are physically equivalent.

Definition: Moving frame. For a given $x \in \mathcal{M}$ and a given space of ‘representative shapes’ $\hat{\mathcal{M}}$, there exists a unique group element $g = g(x, \tau)$ that at instant τ rotates x into $gx = \hat{x} \in \hat{\mathcal{M}}$. The map that associates a state space point x with a group action $g(x, \tau)$ is called a *moving frame*.

exercise A2.1
exercise 13.1

Using decomposition (13.1) one can always write the full state space trajectory as $x(\tau) = g(\tau) \hat{x}(\tau)$, where the $(d-N)$ -dimensional reduced state space trajectory $\hat{x}(\tau)$ is to be fixed by some condition, and $g(\tau)$ is then the corresponding curve on the N -dimensional group manifold of the group action that rotates $\hat{x}$ into x at time τ . The time derivative is then $\dot{x} = v(g\hat{x}) = \dot{g}\hat{x} + g\hat{v}$, with the reduced state space velocity field given by $\hat{v} = d\hat{x}/dt$. Rewriting this as $\hat{v} = g^{-1}v(g\hat{x}) - g^{-1}\dot{g}\hat{x}$ and using the equivariance condition (12.14) leads to

question 13.1

$$\hat{v} = v - g^{-1}\dot{g}\hat{x}. \quad (13.2)$$

The Lie group element (12.5) and its time derivative describe the group tangent flow

$$g^{-1}\dot{g} = g^{-1}\frac{d}{dt}e^{\phi\cdot\mathbf{T}} = \dot{\phi}\cdot\mathbf{T}.$$

This is the group tangent velocity $g^{-1}\dot{g}\hat{x} = \dot{\phi}\cdot t(\hat{x})$ evaluated at the point $\hat{x}$, i.e., with $g = 1$ (see figure 12.3). The flow $\hat{v} = d\hat{x}/dt$ in the $(d-N)$ directions that transverses to the group flow is now obtained by subtracting the flow along the group tangent direction,

$$\hat{v}(\hat{x}) = v(\hat{x}) - \dot{\phi}(\hat{x})\cdot t(\hat{x}). \quad (13.3)$$

We can pick any coordinate transformation (13.1) between the ‘lab’ and a ‘moving frame’, any time, any way we like; equivariance guarantees that the states and the equations of motion (13.3) in the two frames are physically equivalent. This is an immense freedom, and with freedom comes responsibility, the responsibility of choosing a good frame. 

13.2 Symmetry reduction

Maybe when I’m done with grad school I’ll be able to figure it all out . . .

— Rebecca Wilczak, undergraduate

Given Lie group G acting smoothly on a C^∞ manifold $\mathcal{M}$, we can think of each group orbit as an equivalence class. *Symmetry reduction* is the identification of a unique point on each group orbit as the representative of its equivalence class. We call the set of all such group orbit representatives the *reduced state space* $\mathcal{M}/G$. As elaborated in remark 13.1, this space goes by many other names in the literature.

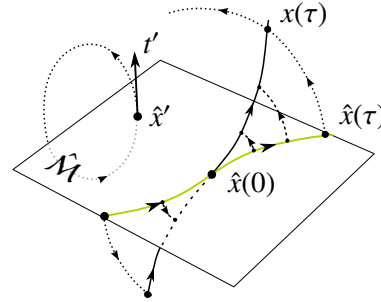
remark 13.1

Symmetry reduction replaces a dynamical system $(\mathcal{M}, f)$ with a symmetry G by a ‘desymmetrized’ system $(\hat{\mathcal{M}}, \hat{f})$ of figure 12.2 (b), a system where each group orbit is replaced by a point, and the action of the group is trivial, $g\hat{x} = \hat{x}$ for all $\hat{x} \in \hat{\mathcal{M}}, g \in G$. The reduced state space $\hat{\mathcal{M}}$ is sometimes called the ‘quotient space’ $\mathcal{M}/G$ because the symmetry has been ‘divided out’. For a discrete symmetry, the reduced state space $\mathcal{M}/G$ is given by the fundamental domain of sect. 11.3. In the presence of a continuous symmetry, the reduction to $\mathcal{M}/G$ amounts to a change of coordinates, where the ‘ignorable angles’ $\{\phi_1, \dots, \phi_N\}$ that parameterize N continuous coordinate transformations are separated out.

13.3 Bringing it all back home: method of slices

In the ‘method of slices’, the reduced state space representative $\hat{x}$ of a group orbit equivalence class is picked by slicing across the group orbits by a fixed hypersurface. 

Figure 13.2: The slice hyperplane $\hat{\mathcal{M}}$, which passes through the template point $\hat{x}'$ and is normal to its group tangent t' , intersects all group orbits (dotted lines) in an open neighborhood of $\hat{x}'$. The full state space trajectory point $x(\tau)$ (solid black line) and the reduced state space trajectory $\hat{x}(\tau)$ (solid green line) belong to the same group orbit $\mathcal{M}_{x(\tau)}$, and are equivalent up to a moving frame group rotation $g(\tau)$ (a rotation by phase $\theta(\tau)$), defined in (13.1).



Definition: Equivariant state space - another word for the ‘full’ state space $\mathcal{M}$. ‘Equivariant’ emphasizes that, in the presence of symmetry, the actions of the symmetry group G stratifies the state space into layers of an onion, or ‘strata’ defined on page 194.

Definition: Reduced state space. A space $\mathcal{M}/G$ in which every group orbit of the equivariant state space $\mathcal{M}$ is represented by a single point.

There are many ways of constructing $\mathcal{M}/G$. One can replace equivariant coordinates $(x_1, x_2, \dots, x_d)$ by a set of invariant polynomials $\{u_1, u_2, \dots, u_m\}$, as in sect. 13.6. Or one can stay in the original state space, but pick a random point on each group orbit and throw away the rest. The most sensible strategy, however, is to smoothly change the coordinates, in such a way that locally the symmetry group acts on N ‘phase’ coordinates and leaves invariant the smooth manifold $\hat{\mathcal{M}} = \mathcal{M}/G$ spanned by the remaining $(d-N)$ transverse coordinates.

Definition: Slice. Consider an open connected d -dimensional domain of the d -dimensional state space manifold $\mathcal{M}$, with all group G -orbits N -dimensional (i.e., the domain excludes any G - or G_p -invariant subspace and its boundary). A *slice* through the *template* point $\hat{x}'$ is a $(d-N)$ -dimensional submanifold $\hat{\mathcal{M}}$ such that all group orbits in an open neighborhood of the point $\hat{x}'$ intersect $\hat{\mathcal{M}}$ transversally once and only once (see figure 13.2).



The simplest *slice condition* defines a linear slice as a $(d-N)$ -dimensional hyperplane $\hat{\mathcal{M}}$ normal to the N group tangents t'_a at template point $\hat{x}'$:

$$\langle \hat{x} - \hat{x}' | t'_a \rangle = 0, \quad t'_a = t_a(\hat{x}') = \mathbf{T}_a \hat{x}'. \quad (13.4)$$

In other words, ‘slice’ is the group orbit analogue of a Poincaré section (3.5) for time-evolution orbits. Each ‘big circle’ –a closed group orbit tangent to t'_a – intersects the hyperplane at least twice (for the very simplest illustration, see example 13.1. As for a Poincaré section (3.4), we add an orientation condition, and select the intersection with the clockwise rotation angle into the slice.



example 13.1
p. 256

As $\langle \hat{x}' | t'_a \rangle = 0$ by the antisymmetry of $\mathbf{T}_a$, the slice condition (13.4) fixes ϕ for a given x by

$$0 = \langle \hat{x} | t'_a \rangle = \langle x | g(\phi)^\top t'_a \rangle, \quad (13.5)$$

where $g^\top$ denotes the transpose of g . The method of moving frames can be interpreted as a change of variables

$$\hat{x}(\tau) = g^{-1}(\tau) x(\tau), \quad (13.6)$$

that is passing to a frame of reference in which condition (13.5) is identically satisfied, see example 13.1. Therefore the name ‘moving frame’. A moving frame should not be confused with the comoving frame, such as the one illustrated in figure 12.8. Each relative equilibrium, relative periodic orbit and general ergodic trajectory has its own comoving frame. In the method of slices one fixes a stationary slice, and rotates all solutions back into the slice.

Moving frames can be utilized in post-processing methods; trajectories are computed in the full state space, then rotated into the slice whenever desired, with the slice condition easily implemented. The slice group tangent t' is a given vector, and $g(\phi)x$ is another vector, linear in x and a function of group parameters ϕ . Rotation parameters ϕ are determined numerically, by a Newton method, through the slice condition (13.5).

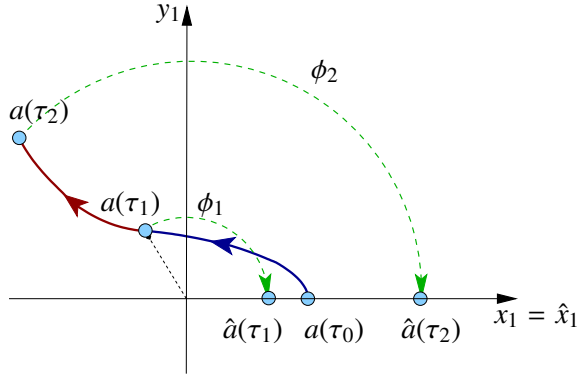
Figure 13.3 illustrates the method of moving frames for an $\text{SO}(2)$ slice normal to the y_1 axis. Looks innocent, but what happens when $(x_1, y_1) = (0, 0)$? More on this in sect. 13.5.

How does one pick a slice point $\hat{x}'$? A generic point $\hat{x}'$ not in an invariant subspace should suffice to fix a slice. The rule of thumb is much like the one for picking Poincaré sections, sect. 3.1.2. This intuitive idea is perhaps best visualized in the context of fluid flows. Suppose the flow exhibits an unstable coherent structure that –approximately– recurs often at different spatial dispositions. One can fit a ‘template’ to one recurrence of such structure and describe other recurrences as its translations. A well chosen slice point belongs to a dynamically important, equivalence class (i.e., group orbit). A slice is locally isomorphic to $\mathcal{M}/G$, in an open neighborhood of $\hat{x}'$. As is the case for the dynamical Poincaré sections, in general a single slice does not suffice to reduce $\mathcal{M} \rightarrow \mathcal{M}/G$ globally.

The Euclidean product of two vectors x, y is indicated in (13.4) by x -transpose times y , as in (12.6). More general bilinear norms $\langle x, y \rangle$ can be used, as long as they are G -invariant, i.e., constant on each irreducible subspace. An example is the quadratic Casimir (12.13).

The slice condition (13.4) fixes N directions, and the remaining vectors $(\hat{x}_{N+1} \dots \hat{x}_d)$ span the slice hyperplane. There are $d-N$ *fundamental invariants*, in the sense that any other invariant can be expressed in terms of them, and they are functionally independent. Thus they serve to distinguish orbits in the neighborhood of the slice-fixing point $\hat{x}'$, i.e., two points lie on the same group orbit if and only if all the fundamental invariants agree.

Figure 13.3: Method of moving frames for the two-modes flow $\text{SO}(2)$ -equivariant under (12.42), with slice through $\hat{a}' = (1, 0, 0, 0)$, group tangent $t' = (0, 1, 0, 0)$. The equivariant flow is 4-dimensional: shown is the projection on the (x_1, y_1) plane. The clockwise orientation condition restricts the slice to the 3-dimensional half-hyperplane $\hat{x}_1 > 0, \hat{y}_1 = 0$. A trajectory started on the slice at $a(\tau_0) = \hat{a}(\tau_0)$, evolves to a state space point with a non-zero $y_1(\tau_n)$. To bring this point back to the slice, compute the polar angle ϕ_n of $a(\tau_n)$ in the (x_1, y_1) plane. Rotate $a(\tau_n)$ clockwise by ϕ_n to $\hat{a}(\tau_n) = g(-\phi_n)a(\tau_n)$, so that the equivalent point on the circle lies on the slice, $\hat{y}_1(\tau_n) = 0$. See sect. 13.5.



13.4 Dynamics within a slice

We made too many wrong mistakes
—Yogi Berra

So far we have taken the post-processing approach: evolve the trajectory in the full state space, then rotate all its points into the slice. You can also split up the time integration into a sequence of finite time steps, each followed by a rotation of the end point into the slice, see figure 13.3. It is tempting to see what happens if the steps are taken infinitesimal. As we shall see, we do get a flow restricted to the slice, but at a price. The relation (13.3) between the ‘lab’ and ‘moving frame’ state space velocity holds for any factorization (13.1) of the flow of form $x(\tau) = g(\tau) \hat{x}(\tau)$. To integrate these equations we first have to fix a particular flow factorization by imposing conditions on $\hat{x}(\tau)$, and then integrate phases $\phi(\tau)$ on a given reduced state space trajectory $\hat{x}(\tau)$.

Here we demand that the reduced state space is confined to a slice hyperplane. Substituting (13.3) into the time derivative of the fixed slice condition (13.5),

$$\langle \hat{v}(\hat{x}) | t'_a \rangle = \langle v(\hat{x}) | t'_a \rangle - \dot{\phi}_b \langle t_b(\hat{x}) | t'_a \rangle = 0,$$

yields the equation for the group phase velocities $\dot{\phi}_a$ for the slice fixed by $\hat{x}'$, together with the reduced state space $\hat{\mathcal{M}}$ flow $\hat{v}(\hat{x})$. In general, the computation of phase velocities requires inversion of the position-dependent $[N \times N]$ matrix $\langle t(\hat{x})_b | t_a \rangle$, so from now on we specialize to the simplest, $N = 1$ parameter case $G = \text{SO}(2)$, where we set $\phi_a = \phi, t'_a = t'$:

$$\hat{v}(\hat{x}) = v(\hat{x}) - \dot{\phi}(\hat{x}) t(\hat{x}), \quad \hat{x} \in \hat{\mathcal{M}} \quad (13.7)$$

$$\dot{\phi}(\hat{x}) = \langle v(\hat{x}) | t' \rangle / \langle t(\hat{x}) | t' \rangle. \quad (13.8)$$

Each group orbit $\mathcal{M}_x = \{g x | g \in G\}$ is an equivalence class; method of slices represents the class by its single slice intersection point $\hat{x}$. By construction $\langle \hat{v} | t' \rangle = 0$, and the motion stays in the $(d-N)$ -dimensional slice. We have thus replaced the original dynamical system $\{\mathcal{M}, f\}$ by a reduced system $\{\hat{\mathcal{M}}, \hat{f}\}$.

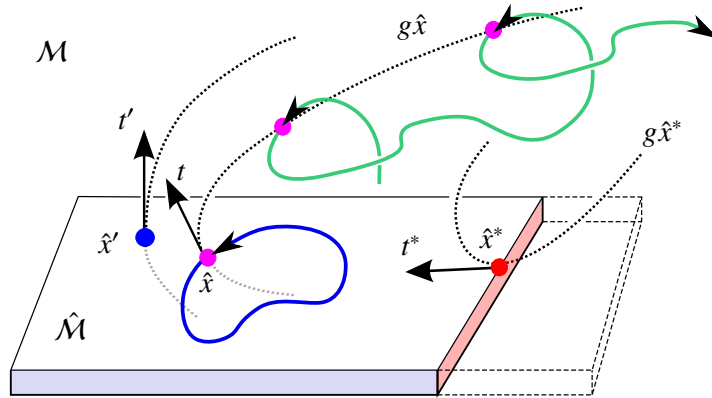


Figure 13.4: Schematic of symmetry reduction by the method of slices. The blue point is the template $\hat{x}'$. All pink points are equivalent to $\hat{x}$ up to a shift, so a relative periodic orbit (green) in the d -dimensional full state space $\mathcal{M}$ closes into a periodic orbit (blue) in the slice $\hat{\mathcal{M}} = \mathcal{M}/G$, a $(d-1)$ -dimensional slab transverse to the template group tangent t' . A typical group orbit (dotted) crosses the slice hyperplane transversally, with a non-orthogonal group tangent $t = t(\hat{x})$. A slice hyperplane is almost never a global slice. It is valid up to the slice border, a $(d-2)$ -dimensional hypersurface (red) of points $\hat{x}^*$ whose group orbits graze the slice, i.e. points whose tangents $t^* = t(\hat{x}^*)$ lie in $\hat{\mathcal{M}}$. Group orbits beyond the slice border do not reach the slice hyperplane: the “missing chunk” is here indicated by the dashed lines.

In the pattern recognition and ‘template fitting’ settings, (13.8) is called the ‘reconstruction equation’. Integrated together, the reduced state space trajectory (13.7) and $g(\tau) = \exp\{\phi(\tau) \cdot \mathbf{T}\}$, the integrated phase (13.8), reconstruct the full state space trajectory $x(\tau) = g(\tau) \hat{x}(\tau)$ from the reduced state space trajectory $\hat{x}(\tau)$, so no information about the flow is lost in the process of symmetry reduction.

exercise 13.2

Slice flow equations (13.7) and (13.8) are pretty, but there is trouble in paradise. The slice flow encounters singularities in subsets of state space, with phase velocity $\dot{\phi}$ divergent whenever the denominator in (13.8) changes sign. We are going to refer to the set of points $\hat{x}^*$ at which the denominator of (13.8) vanishes as *slice border*



$$\langle t(\hat{x}^*) | t' \rangle = 0. \quad (13.9)$$

See figure 13.4 for a schematic illustration. Existence of the slice border makes the method of slices a local method in general, where one constructs a slice by picking a template on a particularly interesting solution and then explores the dynamics nearby. However, this is only partially useful for our purposes, since we would like to explore global objects, such as symmetry-reduced chaotic attractors, interrelation of coherent solutions etc. Several attempts have been made to overcome this problem by defining multiple slices. This is done in such a way that interconnecting the slices does not result in the individual borders of different slices being visited by the dynamics. However, this is a very complicated task and requires case-by-case attention. The other option is defining a very special slice such that its border is not visited by the dynamics. In the next section, we describe such a method which is applicable to many problems that are of interest to us.

13.5 First Fourier mode slice

(N.B. Budanur)

So far, we have given a general description of the method of slices, without specifying the type of symmetries that we are reducing. We have mentioned in sect. 2.4 and sect. 12.1 that the dynamics of nonlinear fields of ‘spatially extended systems’ in periodic cells is a topic of much interest. Such systems are equivariant under spatial translations. Consider a scalar field $u(x, \tau)$ over one space dimension x and time τ , periodic on a spatial domain of length L , $u(x, \tau) = u(x+L, \tau)$. Expand the field in Fourier series, as in (2.19)

$$u(x, \tau) = \sum_{k=-\infty}^{+\infty} \tilde{u}_k(\tau) e^{iq_k x}. \quad (13.10)$$

Spatial translations

$$u(x, \tau) \rightarrow u(x + \ell, \tau) \quad (13.11)$$

correspond to U(1) rotations for the Fourier modes

$$\tilde{u}_k \rightarrow e^{ik\theta} \tilde{u}_k, \text{ where } \theta = 2\pi\ell/L. \quad (13.12)$$

In the state space $a = (x_1, y_1, x_2, y_2, \dots, x_N, y_N)$ spanned by the real and imaginary parts of a finite Fourier mode truncation, $(x_i, y_i) = (\text{Re } \tilde{u}_i, \text{Im } \tilde{u}_i)$, this spatial translation is represented by an SO(2) rotation (see example 12.7)

$$g(\theta) = \begin{pmatrix} R(\theta) & 0 & \cdots & 0 \\ 0 & R(2\theta) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R(N\theta) \end{pmatrix}, \quad (13.13)$$

where $R(n\theta) = \begin{pmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{pmatrix}$,

with the Lie algebra element

$$\mathbf{T} = \begin{pmatrix} 0 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & -2 & \cdots & 0 & 0 \\ 0 & 0 & 2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & -N \\ 0 & 0 & 0 & 0 & \cdots & N & 0 \end{pmatrix}. \quad (13.14)$$

The two-modes system is an example of a system with this kind of symmetry with modes truncated at $N = 2$. We define the first Fourier mode slice as the slice hyperplane in this state space with template

$$\hat{a}' = (1, 0, 0, \dots, 0), \quad (13.15)$$

Figure 13.5: Two-modes flow before (a) and after (b) symmetry reduction by first Fourier mode slice. Here a long trajectory (red and blue) starting on the unstable manifold of the TW_1 (red), until it falls on to the strange attractor (blue) and the shortest relative periodic orbit $\bar{1}$ (magenta). Note that the relative equilibrium becomes an equilibrium, and the relative periodic orbit becomes a periodic orbit after symmetry reduction. (N.B. Budanur)

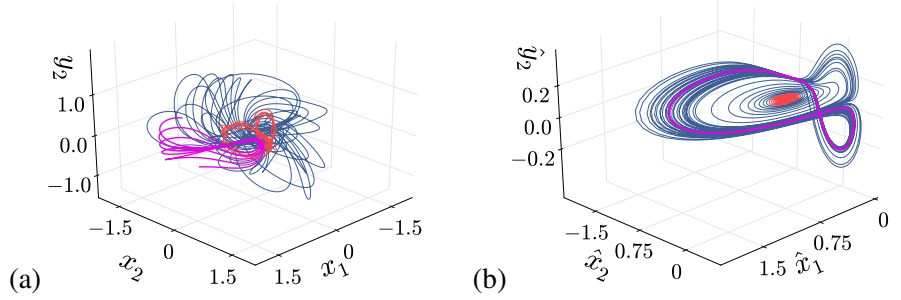
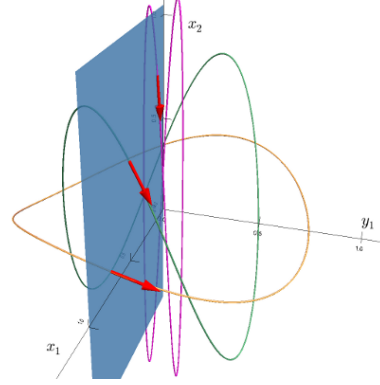


Figure 13.6: $SO(2)$ group orbits of state space points $(0.75, 0, 0.1, 0.1)$ (orange), $(0.5, 0, 0.5, 0.5)$ (green) $(0.1, 0, 0.75, 0.75)$ (magenta) and the first mode slice hyperplane (blue). The group tangents at the intersections with the slice hyperplane are shown as red arrows. As the magnitude of the first Fourier mode decreases relative to the magnitude of the second one, so does the group tangent angle to the slice hyperplane. (from ref. [13]).



and the directional constraint

$$\hat{x}_1 \geq 0 \quad (13.16)$$

(see figure 13.3). We can write equations (13.7) and (13.8), which describe dynamics within the slice hyperplane explicitly for the template (13.15), as

$$\hat{v}(\hat{a}) = v(\hat{a}) - \frac{\dot{y}_1(\hat{a})}{\hat{x}_1} t(\hat{a}), \quad (13.17)$$

$$\dot{\phi}(\hat{a}) = \frac{\dot{y}_1(\hat{a})}{\hat{x}_1}. \quad (13.18)$$

We see from (13.17) and (13.18) that they become singular when $\hat{x}_1 = 0$, i.e. when the amplitude of the first Fourier mode exactly vanishes. In sect. 13.4 we argued that the slice singularity happens when the dot product $t(\hat{a}) \cdot t'$ vanishes, in other words, when the group tangent $t(\hat{a})$ evaluated at the state space point $\hat{a}$ has no component perpendicular to the slice hyperplane. We visualize this in figure 13.6 by showing three dimensional projections of the slice hyperplane, three group orbits and group tangents for the two-modes system.

Our experience from working with spatially extended systems has been that the first Fourier mode amplitude can get very small, but it does not exactly vanish, unless a specific initial condition is set. We can deal with the situations when $\hat{x}_1$ is arbitrarily small by defining the *in-slice time* as

$$d\hat{\tau} = d\tau / \hat{x}_1 \quad (13.19)$$

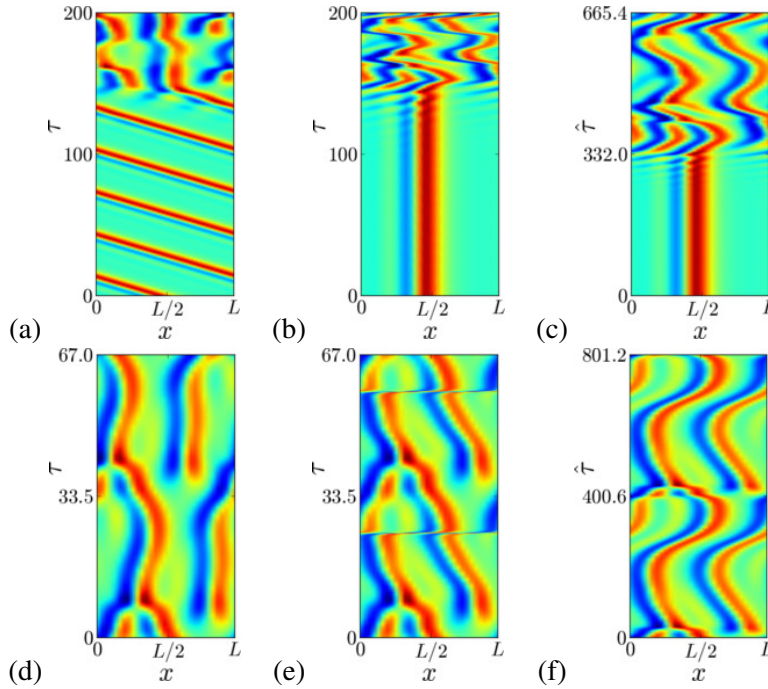


Figure 13.7: Traveling wave TW_1 with phase velocity $c = 0.737$ in configuration space: (a) the full state space solution, (b) symmetry-reduced solution with respect to the lab time, and (c) symmetry-reduced solution with respect to the in-slice time. Two repeats of $T = 33.50$ relative periodic orbit in configuration space: (d) the full state space solution, (e) symmetry-reduced solution with respect to the lab time, and (f) symmetry-reduced solution with respect to the in-slice time (from ref. [14]).

and re-writing (13.17) and (13.18) in terms of $d\hat{\tau}$ as

$$d\hat{a}/d\hat{\tau} = \hat{x}_1 v(\hat{a}) - \hat{y}_1(\hat{a}) t(\hat{a}), \quad (13.20)$$

$$d\theta(\hat{a})/d\hat{\tau} = \hat{y}_1(\hat{a}). \quad (13.21)$$

One is assured a smooth flow by integrating (13.20) to obtain symmetry-invariant dynamics. Figure 13.7 illustrates the importance of time rescaling on the first Fourier mode slice to the Kuramoto-Sivashinsky system.

 example 13.2
p. 256

13.6 Method of images: Hilbert bases

(E. Siminos and P. Cvitanović)

An erudite reader might wonder: why all this slicing and dicing, when the problem of symmetry reduction was solved by Hilbert and Weyl a century ago? Indeed, the most common approach to symmetry reduction is by means of a Hilbert invariant polynomial basis (11.7), motivated intuitively by the existence of such nonlinear invariants as the rotationally-invariant length $r^2 = x_1^2 + x_2^2 + \dots + x_d^2$,

or, in Hamiltonian dynamics, the energy function. One trades in the equivariant state space coordinates $\{x_1, x_2, \dots, x_d\}$ for a non-unique set of $m \geq d$ polynomials $\{u_1, u_2, \dots, u_m\}$ invariant under the action of the symmetry group. These polynomials are linearly independent, but functionally dependent through $m - d + N$ relations called *syzygies*.

The dynamical equations follow from the chain rule

$$\dot{u}_i = \frac{\partial u_i}{\partial x_j} \dot{x}_j, \quad (13.22)$$

upon substitution $\{x_1, x_2, \dots, x_d\} \rightarrow \{u_1, u_2, \dots, u_m\}$. One can either rewrite the dynamics in this basis or plot the ‘image’ of solutions computed in the original, equivariant basis in terms of these invariant polynomials.

Nevertheless we can now easily identify a suitable Poincaré section, guided by the Lorenz flow examples of chapter 11, as one that contains the z -axis and the image of the relative equilibrium TW_1 , here defined by the condition $u_1 = u_4$. As in example 14.4, we construct the first return map using as coordinate the Euclidean arclength along the intersection of the unstable manifold of TW_1 with the Poincaré section. Thus the goals set in the introduction to this chapter are attained. We have reduced the messy strange attractor of figure 12.1 to a 1-dimensional return map. We now have the symbolic dynamics and can compute as many relative periodic orbits of the complex Lorenz flow as we wish, missing none. This will be explained further, when we discuss the Lorenz attractor in example 14.4.

Reducing dimensionality of a dynamical system by explicit elimination of variables through inclusion of syzygies introduces singularities. Such elimination of variables, however, is not needed for visualization purposes; syzygies merely guarantee that the dynamics takes place on a $(d - N)$ -dimensional submanifold in the projection on the $\{u_1, u_2, \dots, u_m\}$ coordinates. However, when one *reconstructs* the dynamics in the original space $\mathcal{M}$ from its image $\mathcal{M}/G$, the transformations have singularities at the fixed-point subspaces of the isotropy subgroups in $\mathcal{M}$.

What limits the utility of Hilbert basis methods are not such singularities, but rather the fact that the algebra needed to determine a Hilbert basis becomes computationally prohibitive as the dimension of the system and/or the symmetry group increases. Moreover, even if such basis were available, rewriting the equations in an invariant polynomial basis seems impractical, so in practice Hilbert basis computations appear not feasible beyond state space dimension of order $\approx$ ten. When the goal is to quotient continuous symmetries of high-dimensional flows, such as the Navier-Stokes flows, one needs a workable framework. The method of slices of sect. 13.2 is one such minimalist alternative.

Résumé

Here we have described how, and offered two approaches to continuous symmetry reduction. In the *method of slices* one fixes a ‘slice’ $\langle \hat{x} - \hat{x}' | t' \rangle = 0$, a hyperplane normal to the group tangent t' , that cuts across group orbits in the neighborhood of the slice-fixing point $\hat{x}'$. Each class of symmetry-equivalent points is represented by a single point, with the symmetry-reduced dynamics in the reduced state space $\mathcal{M}/G$ given by (13.7):

$$\hat{v} = v - \dot{\phi} \cdot t, \quad \dot{\phi} = \langle v | t' \rangle / \langle t | t' \rangle.$$

In practice one runs the dynamics in the full state space, and post-processes the trajectory by the method of moving frames. In the *Hilbert polynomial basis* approach one transforms the equivariant state space coordinates into invariant ones, by a nonlinear coordinate transformation

$$\{x_1, x_2, \dots, x_d\} \rightarrow \{u_1, u_2, \dots, u_m\} + \{\text{syzygies}\},$$

and studies the invariant ‘image’ of dynamics (13.22) rewritten in terms of invariant coordinates.

Continuous symmetry reduction is considerably more involved than the discrete symmetry reduction to a fundamental domain of chapter 11. Slices are only local sections of group orbits, and Hilbert polynomials are non-unique and difficult to compute for high-dimensional flows. However, there is no need to actually recast the dynamics in the new coordinates. Either approach can be used as a visualization tool, with all computations carried out in the original coordinates, and then, when done, rotating the solutions into the symmetry reduced state space by post-processing the data. The trick is to construct a good set of symmetry invariant Poincaré sections (see sect. 3.1), and that is always a dark art, with or without a symmetry.

Relative equilibria and relative periodic orbits are the hallmark of systems with continuous symmetry. Amusingly, in this extension of ‘periodic orbit’ theory from unstable 1-dimensional closed periodic orbits to unstable $(N+1)$ -dimensional compact manifolds $\mathcal{M}_p$ (invariant under continuous symmetries), either relatively few or no periodic orbits exist. Relative periodic orbits are almost never eventually periodic, i.e., they almost never lie on periodic trajectories in the full state space, so looking for periodic orbits in systems with continuous symmetries is a fool’s errand.

However, dynamical systems are often equivariant under a combination of continuous symmetries and discrete coordinate transformations of chapter 10. An example is the orthogonal group $O(n)$. In the presence of discrete symmetries, relative periodic orbits within discrete symmetry-invariant subspaces are eventually periodic. Atypical as they are (no generic chaotic orbit can ever enter these discrete invariant subspaces), they tend to be the most symmetric solutions. Additionally, these relative periodic orbits will be important for periodic orbit theory, where the shortest orbits dominate.

The message: If a dynamical systems has a symmetry, use it!

chapter 25

Commentary

Remark 13.1. A brief history of relativity, or, ‘Desymmetrization and its discontents’ (after [Civilization and its discontents](#); continued from remark [12.1](#)).

In ChaosBook ‘symmetry reduction’ is the identification of a unique point on a group orbit as the representative of its equivalence class. We call the set of all such group orbit representatives the *reduced state space* M/G . This space has many names in the literature: ‘desymmetrized state space’, ‘symmetry-reduced space’, ‘orbit space’ (because every group orbit in the original space is represented by a single point in the orbit space), ‘base manifold’, ‘shape-changing space’ or ‘quotient space’, obtained by mapping equivariant dynamics to invariant dynamics (‘image’) by methods such as ‘moving frames’, ‘cross sections’, ‘slices’, ‘freezing’, ‘Hilbert bases’, ‘quotienting’, ‘lowering of the degree’, ‘lowering the order’, or ‘desymmetrization’.

Relative equilibria and relative periodic solutions are related by symmetry reduction to equilibria and periodic solutions of the reduced dynamics. They appear in many physical applications, such as celestial mechanics, molecular dynamics, motion of rigid bodies, nonlinear waves, spiralling patterns, and fluid mechanics. A relative equilibrium is a solution which travels along an orbit of the symmetry group at constant speed; one good introduction to these relative equilibria can be found in Marsden [\[46\]](#). According to Cushman, Bates [\[21\]](#) and Yoder [\[79\]](#), C. Huygens [\[39\]](#) understood the relative equilibria of a spherical pendulum many years before publishing them in 1673. A reduction of the translation symmetry was obtained by Jacobi (see [wiki/Jacobi coordinates](#); for a modern, symplectic implementation, see Laskar *et al.* [\[45\]](#)). In 1892 German sociologist Vierkandt [\[73\]](#) showed that on a symmetry-reduced space (the constrained velocity phase space modulo the action of the group of Euclidean motions of the plane) all orbits of the rolling disk system are periodic [\[8\]](#). According to Chenciner [\[16\]](#), the first attempt to find (relative) periodic solutions of the N -body problem was the 1896 short note by Poincaré [\[59\]](#), in the context of the 3-body problem. Poincaré named such solutions ‘relative’. Relative equilibria of the N -body problem (known in this context as the Lagrange points, stationary in the co-rotating frame) are circular motions in the inertial frame, and relative periodic orbits correspond to quasiperiodic motions in the inertial frame. For relative periodic orbits in celestial mechanics see also ref. [\[12\]](#). A striking application of relative periodic orbits has been the discovery of “choreographies” in the N -body problems [\[17, 18, 49\]](#).

The modern story on equivariance and dynamical systems starts perhaps with S. Smale [\[69\]](#), M. Field [\[27\]](#), and bifurcations in the presence of symmetries with Ruelle [\[63\]](#). Ruelle proves that (1) the stability matrix/Jacobian matrix evaluated at an equilibrium/fixed point $x \in M_G$ decomposes into linear irreps of G , (2) stable/unstable manifold continuations of its eigenvectors inherit their symmetry properties, and (3) an equilibrium can bifurcate to a rotationally invariant periodic orbit (i.e., relative equilibrium).

Gilmore and Lettelier monograph [\[33\]](#) offers a very clear, detailed and user friendly discussion of symmetry reduction by means of Hilbert polynomial bases (do not look for ‘Hilbert’ in the index, though). Vladimirov, Toronov and Derbov [\[75\]](#) use an invariant polynomial basis to study bounding manifolds of the symmetry reduced complex Lorenz flow and its homoclinic bifurcations. There is no general strategy how to construct a Hilbert basis; clever low-dimensional examples have been constructed case-by-case. The determination of a Hilbert basis appears computationally prohibitive for state space dimensions larger than ten [\[19, 29\]](#), and rewriting the equations of motions in invariant polynomial bases appears impractical for high-dimensional flows.

Hilbert proved [\[38\]](#) the theorem [\(11.7\)](#) for the group $SL(2, \mathbb{C})$ and Emily Noether [\[55\]](#)

proved it for all the finite groups. Thus, by 1920's the problem of rewriting equivariant flows as invariant ones was solved by Hilbert and Weyl, but at the cost of introducing largely arbitrary extra dimensions, with the reduced flows on manifolds of lower dimensions, constrained by sets of syzygies. Cartan found this unsatisfactory, and in 1935 he introduced [15] the notion of a *moving frame*, a map from a manifold to a Lie group. This mapping seeks no invariant polynomial basis, but instead rewrites the reduced M/G flow in terms of $d-N$ *fundamental invariants* defined by a *slice* that cuts across all group orbits in some open neighborhood. Fels and Olver view the method as an alternative to the Gröbner bases methods for computing Hilbert polynomials, to compute functionally independent fundamental invariant bases for general group actions (with no explicit connection to dynamics, differential equations or symmetry reduction). 'Fundamental' here means that they can be used to generate all other invariants. Olver's monograph [57] is pedagogical, but does not describe the original Cartan's method. Fels and Olver papers [23, 24] are lengthy and technical. They refer to Cartan's method as a method of 'moving frames' and view it as a special and less rigorous case of their 'moving coframe' method. The name 'moving coframes' arises through the use of Maurer-Cartan form which is a coframe on the Lie group G , i.e., they form a pointwise basis for the cotangent space. In refs. [66, 67] the invariant bases generated by the moving frame method are used as a basis to project a full state space trajectory to the slice (i.e., the M/G reduced state space).

The basic idea of the 'method of slices' is intuitive and frequently reinvented, often under a different name; for example, it is stated without attribution as in problem 1 of Sect. 6.2 of Arnol'd *Ordinary Differential Equations* [2]. The factorization (13.1) is stated on p. 31 of Anosov and Arnol'd [1], who note, without further elaboration, that in the vicinity of a point which is not fixed by the group one can reduce the order of a system of differential equations by the dimension of the group. Ref. [3] refers to symmetry reduction as 'lowering the order'. For the definition of 'slice' see, for example, Chossat and Lauterbach [19]. Briefly, a submanifold $M_{\hat{x}}$ containing $\hat{x}$ is called a *slice* through $\hat{x}$ if it is invariant under isotropy $G_{\hat{x}}(M_{\hat{x}}) = M_{\hat{x}}$. If $\hat{x}$ is a fixed point of G , then slice is invariant under the whole group. One good description of slice theorem can be found in [Encyclopaedia of Mathematics](#). Slices tend to be discussed in contexts more difficult than our application - symplectic groups, sections in absence of global charts, non-compact Lie groups. We follow ref. [62] in referring to a local group-orbit section as a 'slice'. Refs. [11, 34] and others refer to global group-orbit sections as 'cross-sections', a term that we would rather avoid, as it already has a different and well established meaning in physics. Duistermaat and Kolk [22] refer to 'slices', but the usage goes back at least to Guillemin and Sternberg [34] in 1984, Palais [58] in 1961 and Mostow [51] in 1957 (who discusses "local cross-sections"). Bredon [11] discusses both cross-sections and slices. Guillemin and Sternberg [34] define the 'cross-section', but emphasize that finding it is very rare: "existence of a global section is a very stringent condition on a group action. The notion of 'slice' is weaker but has a much broader range of existence."

Several important fluid dynamics flows exhibit continuous symmetries which are either $SO(2)$ or products of $SO(2)$ groups, each of which act on different coordinates of the state space. The Kuramoto-Sivashinsky equations [43, 68], plane Couette flow [30, 31, 35, 74], and pipe flow [40, 76] all have continuous symmetries of this form. In the 1982 paper Rand [60] explains how the presence of continuous symmetries gives rise to rotating and modulated rotating (quasiperiodic) waves in fluid dynamics. Haller and Mezić [36] reduce symmetries of three-dimensional volume-preserving flows and reinvent method of moving frames, under the name 'orbit projection map'. There is extensive literature on reduction of symplectic manifolds with symmetry; Marsden and Weinstein 1974 article [48] is an important early reference. Then there are studies of the reduced phase spaces for vortices moving on a sphere such as ref. [41], and many, many others.

Reaction-diffusion systems are often equivariant with respect to the action of a finite

dimensional (not necessarily compact) Lie group. Spiral wave formation in such nonlinear excitable media was first observed in 1970 by Zaikin and Zhabotinsky [80]. Winfree [77, 78] noted that spiral tips execute meandering motions. Barkley and collaborators [4, 5] showed that the noncompact Euclidean symmetry of this class of systems precludes nonlinear entrainment of translational and rotational drifts and that the interaction of the Hopf and the Euclidean eigenmodes leads to observed quasiperiodic and meandering behaviors. Fiedler, in the influential 1995 talk at the Newton Institute, and Fiedler, Sandstede, Wulff, Turaev and Scheel [25, 26, 64, 65] treat Euclidean symmetry bifurcations in the context of spiral wave formation. The central idea is to utilize the semidirect product structure of the Euclidean group $E(2)$ to transform the flow into a ‘skew product’ form, with a part orthogonal to the group orbit, and the other part within it, as in (13.7). They refer to a linear slice $\hat{\mathcal{M}}$ near relative equilibrium as a *Palais slice*, with Palais coordinates. As the choice of the slice is arbitrary, these coordinates are not unique. According to these authors, the skew product flow was first written down by Mielke [50], in the context of buckling in elasticity theory. However, this decomposition is no doubt much older. For example, it was used by Krupa [19, 42] in his local slice study of bifurcations of relative equilibria. Biktashev, Holden, and Nikolaev [7] cite Anosov and Arnol’d [1] for the ‘well-known’ factorization (13.1) and write down the slice flow equations (13.7).

Neither Fiedler *et al.* [25] nor Biktashev *et al.* [7] implemented their methods numerically. That was done by Rowley and Marsden for the Kuramoto-Sivashinsky [62] and Burgers [61] equations, and Beyn and Thümmmler [6, 70] for a number of reaction-diffusion systems, described by parabolic partial differential equations on unbounded domains. We recommend the Barkley paper [4] for a clear explanation of how the Euclidean symmetry leads to spirals, and the Beyn and Thümmmler paper [6] for inspirational concrete examples of how ‘freezing’/‘slicing’ simplifies the dynamics of rotational, traveling and spiraling relative equilibria. Beyn and Thümmmler write the solution as a composition of the action of a time dependent group element $g(\tau)$ with a ‘frozen’, in-slice solution $\hat{u}(\tau)$ (13.1). In their nomenclature, making a relative equilibrium stationary by going to a comoving frame is ‘freezing’ the traveling wave, and the imposition of the phase condition (i.e., slice condition (13.4)) is the ‘freezing ansatz’. They find it more convenient to make use of the equivariance by extending the state space rather than reducing it. This technique adds an additional parameter and a phase condition. The ‘freezing ansatz’ [6] is identical to Rowley and Marsden [61] and our slicing, except that ‘freezing’ is formulated as an additional constraint, just as when we compute periodic orbits of ODEs we add a Poincaré section as an additional constraint, i.e., increase the dimensionality of the problem by 1 for every continuous symmetry (see sect. 7.2).

section 7.2

Several symmetry reduction schemes are reviewed in ref. [67]. Here we describe the method of slices [6, 28, 62], the only method that we find practical for a symmetry reduction of chaotic solutions of highly nonlinear and possibly also high-dimensional flows. Derivation of sect. 13.4 follows most closely Rowley and Marsden [61] who, in the pattern recognition setting refer to the slice point as a ‘template’, and call (13.8) the ‘reconstruction equation’ [46, 47]. They also describe the ‘method of connections’ (called ‘orthogonality of time and group orbit at successive times’ in ref. [6]), for which the reconstruction equation (13.8) denominator is $\langle t(\hat{x})|t(\hat{x}) \rangle$ and thus non-vanishing as long as the action of the group is regular. This avoids spurious slice singularities, but it is not clear what the ‘method of connections’ buys us otherwise. It does not reduce the dimensionality of the state space, and it accrues ‘geometric phases’ which prevent relative periodic orbits from closing into periodic orbits. The geometric phase in laser equations, including complex Lorenz equations, has been studied in ref. [52–54, 71, 72]. Another theorist’s temptation is to hope that a continuous symmetry would lead us to a conserved quantity. However, the Noether theorem [56] requires that equations of motion be cast in Lagrangian form and that the Lagrangian exhibits variational symmetries [9, 10]. Such variational symmetries are hard to find for dissipative systems.

In general relativity ‘symmetry reduction’ is a method of finding exact solutions by *imposing* symmetry conditions to obtain a reduced system of equations, i.e., restricting the set of solutions considered to an invariant subspace. This is absolutely not what we mean by ‘symmetry reduction’ in ChaosBook. In Quantum Field Theory van den Heuvel and van Baal [37], for example, mean slice when they say “fundamental domain”: “Let $\mathcal{A}$ be the set of all gauge fields $\mathcal{A} : S^3 \rightarrow su(2)$. The *physical configuration space* is the space of gauge orbits $\mathcal{A}/G$. Let G be the set of all gauge transformations $g : S^3 \rightarrow SU(2)$. We would like to have a *fundamental domain*, that is, a set of gauge fields which is in one-to-one correspondence with the physical configuration space.” This work is continued, for example, by Cucchieri [20] who studies the so-called *fundamental modular region*. This region is free of Gribov copies, in the minimal Landau gauge for pure $SU(2)$ lattice gauge theory.

References to ‘cyclists’ are bit of a joke in more ways than one. First, ‘cyclist’, ‘pedestrian’ throughout ChaosBook.org refer jokingly both to the title of Lipkin’s *Lie Groups for Pedestrians* [44] and to our preoccupations with [actual cycling](#). Lipkin’s ‘pedestrian’ is fluent in Quantum Field Theory, but wobbly on Dynkin diagrams. More to the point, it is impossible to dispose of Lie groups in a page of text. As an antidote to the brevity of exposition here, consider reading Gilmore’s monograph [32] which offers a quirky, personal and enjoyable distillation of a lifetime of pondering Lie groups. As seems to be the case with any textbook on Lie groups, it will not help you with the problem at hand, but it is the only place you can learn both what Galois actually did when he invented the theory of finite groups in 1830, and what, inspired by Galois, Lie actually did in his 1874 study of symmetries of ODEs. Gilmore also explains many things that we pass over in silence here, such as matrix groups, group manifolds, and compact groups.

One would think that with all this literature the case is shut and closed, but not so. Applied mathematicians are inordinately fond of bifurcations, and almost all of the published work focuses on equilibria, relative equilibria, and their bifurcations, and for these problems a single slice works well. Only when one tries to describe the totality of chaotic orbits does the non-global nature of slices become a serious nuisance.

(E. Siminos and P. Cvitanović)

Question 13.1. Alice is not happy

Q All these formulas like $\hat{v} = v - g^{-1} \dot{g} \hat{x}$ look terribly formal...

A Not at all - you use them all the time. What in this classroom is a left, a right, up, down is actually a wildly rotating frame, whirling along with the rotation of our globe, which is rotating around the sun, and so on. If you only care about what happens in this room, $\hat{v}(\tau)$ is a good description of the evolution law. If you want to compare celestial observations from different telescopes, you better go with $g(\tau)$ into the ‘fixed’ celestial frame, where the evolution law is of form $v(\tau)$.

Question 13.2. Alice asks

Q What is the difference between ‘velocity’ and ‘speed’?

A *Velocity* is a vector, the rate at which the object changes its position. *Speed*, or the magnitude of the velocity, is a scalar quantity which describes how fast an object moves. We denote the rate of change of group phases, or the *phase velocity* by the vector $c(x) = (\dot{\phi}_1, \dots, \dot{\phi}_N) = (c_1, \dots, c_N)$, a component for each of the N continuous symmetry parameters. These are converted to state space velocity components along the group tangents by

$$v_t(x) = c(x) \cdot t(x). \quad (13.23)$$

For rotational waves these are called ‘angular velocities’.

Remark 13.2. Killing fields. The symmetry tangent vector fields discussed here are a special case of Killing vector fields of Riemannian geometry and special relativity. If this poetry warms the cockles of your heart, hang on. From [wikipedia](#) ([this wikipedia](#) might also be useful): A Killing vector field is a set of infinitesimal generators of isometries on a Riemannian manifold that preserve the metric. Flows generated by Killing fields are continuous isometries of the manifold. The flow generates a symmetry, in the sense that moving each point on an object the same distance in the direction of the Killing vector field will not distort distances on the object. A vector field X is a Killing field if the Lie derivative with respect to X of the metric g vanishes:

$$\mathcal{L}_X g = 0. \quad (13.24)$$

Killing vector fields can also be defined on any (possibly nonmetric) manifold $\mathcal{M}$ if we take any Lie group G acting on it instead of the group of isometries. In this broader sense, a Killing vector field is the pushforward of a left invariant vector field on G by the group action. The space of the Killing vector fields is isomorphic to the Lie algebra $\mathfrak{g}$ of G .

If the equations of motion can be cast in Lagrangian form, with the Lagrangian exhibiting variational symmetries [9, 10], Noether theorem associates a conserved quantity with each Killing vector.

References

- [1] D. V. Anosov and V. I. Arnol'd, *Dynamical systems I: Ordinary Differential Equations and Smooth Dynamical Systems* (Springer, New York, 1988).
- [2] V. I. Arnol'd, *Ordinary Differential Equations* (Springer, New York, 1992).
- [3] V. I. Arnol'd, V. V. Kozlov, and A. I. Neishtadt, *Mathematical Aspects of Classical and Celestial Mechanics* (Springer, New York, 1988).
- [4] D. Barkley, "Euclidean symmetry and the dynamics of rotating spiral waves", *Phys. Rev. Lett.* **72**, 164–167 (1994).
- [5] D. Barkley, M. Kness, and L. S. Tuckerman, "Spiral wave dynamics in a simple model of excitable media: Transition from simple to compound rotation", *Phys. Rev. A* **42**, 2489–2492 (1990).
- [6] W.-J. Beyn and V. Thümmel, "Freezing solutions of equivariant evolution equations", *SIAM J. Appl. Dyn. Syst.* **3**, 85–116 (2004).
- [7] V. N. Biktashev, A. V. Holden, and E. V. Nikolaev, "Spiral wave meander and symmetry of the plane", *Int. J. Bifur. Chaos* **6**, 2433–2440 (1996).
- [8] A. M. Bloch, J. E. Marsden, and D. V. Zenkov, "Nonholonomic dynamics", *Notices Amer. Math. Soc.* **52**, 324–333 (2005).
- [9] G. Bluman, "Connections between symmetries and conservation laws", *SIGMA* **1**, 011 (2005).
- [10] G. W. Bluman and S. C. Anco, *Symmetry and Integration Methods for Differential Equations* (Springer, New York, 2002).
- [11] G. Bredon, *Introduction to Compact Transformation Groups* (Academic, New York, 1972).

- [12] R. Broucke, “On relative periodic solutions of the planar general three-body problem”, *Celestial Mech. Dynam. Astronom.* **12**, 439–462 (1975).
- [13] N. B. Budanur, D. Borrero-Echeverry, and P. Cvitanović, “Periodic orbit analysis of a system with continuous symmetry - A tutorial”, *Chaos* **25**, 073112 (2015).
- [14] N. B. Budanur, P. Cvitanović, R. L. Davidchack, and E. Siminos, “Reduction of the $SO(2)$ symmetry for spatially extended dynamical systems”, *Phys. Rev. Lett.* **114**, 084102 (2015).
- [15] É. Cartan, *La Méthode du Repère Mobile, la Théorie des Groupes Continus, et les Espaces Généralisés*, Vol. 5, Exposés de Géométrie (Hermann, Paris, 1935).
- [16] A. Chenciner, “A note by Poincaré”, *Regul. Chaotic Dyn.* **10**, 119–128 (2005).
- [17] A. Chenciner, J. Gerver, R. Montgomery, and C. Simó, “Simple choreographic motions of N -bodies: A preliminary study”, in *Geometry, mechanics and dynamics*, edited by P. Newton, P. Holmes, and A. Weinstein (Springer, New York, 2002), pp. 287–308.
- [18] A. Chenciner and R. Montgomery, “A remarkable solution of the 3-body problem in the case of equal masses”, *Ann. Math.* **152**, 881–901 (2000).
- [19] P. Chossat and R. Lauterbach, *Methods in Equivariant Bifurcations and Dynamical Systems* (World Scientific, Singapore, 2000).
- [20] A. Cucchieri, “Numerical study of the fundamental modular region in the minimal Landau gauge”, *Nucl. Phys. B* **521**, 365–379 (1998).
- [21] R. H. Cushman and L. M. Bates, *Global Aspects of Classical Integrable Systems* (Birkhäuser, Boston, 1997).
- [22] J. J. Duistermaat and J. A. C. Kolk, *Lie Groups* (Springer, New York, 2000).
- [23] M. Fels and P. J. Olver, “Moving coframes: I. A practical algorithm”, *Acta Appl. Math.* **51**, 161–213 (1998).
- [24] M. Fels and P. J. Olver, “Moving coframes: II. Regularization and theoretical foundations”, *Acta Appl. Math.* **55**, 127–208 (1999).
- [25] B. Fiedler, B. Sandstede, A. Scheel, and C. Wulff, “Bifurcation from relative equilibria of noncompact group actions: Skew products, meanders, and drifts”, *Doc. Math.* **141**, 479–505 (1996).
- [26] B. Fiedler and D. Turaev, “Normal forms, resonances, and meandering tip motions near relative equilibria of Euclidean group actions”, *Arch. Rational Mech. Anal.* **145**, 129–159 (1998).
- [27] M. Field, “Equivariant dynamical systems”, *Bull. Amer. Math. Soc.* **76**, 1314–1318 (1970).
- [28] S. Froehlich and P. Cvitanović, “Reduction of continuous symmetries of chaotic flows by the method of slices”, *Commun. Nonlinear Sci. Numer. Simul.* **17**, 2074–2084 (2012).

- [29] K. Gatermann, *Computer Algebra Methods for Equivariant Dynamical Systems* (Springer, New York, 2000).
- [30] J. F. Gibson, J. Halcrow, and P. Cvitanović, “Visualizing the geometry of state-space in plane Couette flow”, *J. Fluid Mech.* **611**, 107–130 (2008).
- [31] J. F. Gibson, J. Halcrow, and P. Cvitanović, “Equilibrium and traveling-wave solutions of plane Couette flow”, *J. Fluid Mech.* **638**, 243–266 (2009).
- [32] R. Gilmore, *Lie Groups, Physics, and Geometry* (Cambridge Univ. Press, Cambridge UK, 2008).
- [33] R. Gilmore and C. Letellier, *The Symmetry of Chaos* (Oxford Univ. Press, Oxford, 2007).
- [34] V. Guillemin and S. Sternberg, *Symplectic Techniques in Physics* (Cambridge Univ. Press, Cambridge UK, 1990).
- [35] J. Halcrow, *Geometry of Turbulence: An Exploration of the State-space of Plane Couette Flow*, PhD thesis (School of Physics, Georgia Inst. of Technology, Atlanta, 2008).
- [36] G. Haller and I. Mezić, “Reduction of three-dimensional, volume-preserving flows with symmetry”, *Nonlinearity* **11**, 319–339 (1998).
- [37] B. van den Heuvel and P. van Baal, “Dynamics on the SU(2) fundamental domain”, *Nucl. Phys. Proc. Suppl.* **42**, 823–825 (1995).
- [38] D. Hilbert, “Über die vollen Invariantensysteme”, *Math. Ann.* **42**, 313–373 (1893).
- [39] C. Huygens, *L’Horloge à Pendule* (Swets & Zeitlinger, Amsterdam, 1673).
- [40] R. R. Kerswell, “Recent progress in understanding the transition to turbulence in a pipe”, *Nonlinearity* **18**, R17–R44 (2005).
- [41] F. Kirwan, “The topology of reduced phase spaces of the motion of vortices on a sphere”, *Physica D* **30**, 99–123 (1988).
- [42] M. Krupa, “Bifurcations of relative equilibria”, *SIAM J. Math. Anal.* **21**, 1453–1486 (1990).
- [43] Y. Kuramoto and T. Tsuzuki, “Persistent propagation of concentration waves in dissipative media far from thermal equilibrium”, *Progr. Theor. Phys.* **55**, 365–369 (1976).
- [44] H. J. Lipkin, *Lie Groups for Pedestrians* (North-Holland, Amsterdam, 1966).
- [45] F. Malige, P. Robutel, and J. Laskar, “Partial reduction in the N-body planetary problem using the angular momentum integral”, *Celestial Mech. Dynam. Astronom.* **84**, 283–316 (2002).
- [46] J. E. Marsden, *Lectures on Mechanics* (Cambridge Univ. Press, Cambridge UK, 1992).
- [47] J. E. Marsden and T. S. Ratiu, *Introduction to Mechanics and Symmetry* (Springer, New York, 1999).
- [48] J. E. Marsden and A. Weinstein, “Reduction of symplectic manifolds with symmetry”, *Rep. Math. Phys.* **5**, 121–30 (1974).

- [49] C. McCord, J. Montaldi, M. Roberts, and L. Sbano, [Relative periodic orbits of symmetric Lagrangian systems](#), in *Proc. Equadiff 2003*, edited by F. Dumortier and et.al. (2004), pp. 482–493.
- [50] A. Mielke, *Hamiltonian and Lagrangian Flows on Center Manifolds* (Springer, New York, 1991).
- [51] G. D. Mostow, [“Equivariant embeddings in Euclidean space”](#), *Ann. Math.* **65**, 432–446 (1957).
- [52] C. Z. Ning and H. Haken, [“Phase anholonomy in dissipative optical-systems with periodic oscillations”](#), *Phys. Rev. A* **43**, 6410–6413 (1991).
- [53] C. Z. Ning and H. Haken, [“An invariance property of the geometrical phase and its consequence in detuned lasers”](#), *Z. Phys. B* **89**, 261–262 (1992).
- [54] C. Z. Ning and H. Haken, [“Geometrical phase and amplitude accumulations in dissipative systems with cyclic attractors”](#), *Phys. Rev. Lett.* **68**, 2109–2112 (1992).
- [55] E. Noether, [“Der Endlichkeitssatz der Invarianten endlicher Gruppen”](#), *Math. Ann.* **77**, 89–92 (1915).
- [56] E. Noether, [“Invariante Variationsprobleme”](#), *Nachr. Ges. Wiss. Göttingen, Math.-Phys.* **1918**, 235–257 (1918).
- [57] P. J. Olver, *Classical Invariant Theory* (Cambridge Univ. Press, Cambridge UK, 1999).
- [58] R. S. Palais, [“On the existence of slices for actions of non-compact Lie groups”](#), *Ann. Math.* **73**, 295–323 (1961).
- [59] H. Poincaré, [“Sur les solutions périodiques et le principe de moindre action”](#), *C. R. Acad. Sci. Paris* **123**, 915–918 (1896).
- [60] D. Rand, [“Dynamics and symmetry. Predictions for modulated waves in rotating fluids”](#), *Arch. Rational Mech. Anal.* **79**, 1–37 (1982).
- [61] C. W. Rowley, I. G. Kevrekidis, J. E. Marsden, and K. Lust, [“Reduction and reconstruction for self-similar dynamical systems”](#), *Nonlinearity* **16**, 1257–1275 (2003).
- [62] C. W. Rowley and J. E. Marsden, [“Reconstruction equations and the Karhunen-Loève expansion for systems with symmetry”](#), *Physica D* **142**, 1–19 (2000).
- [63] D. Ruelle, [“Bifurcations in presence of a symmetry group”](#), *Arch. Rational Mech. Anal.* **51**, 136–152 (1973).
- [64] B. Sandstede, A. Scheel, and C. Wulff, [“Dynamics of spiral waves on unbounded domains using center-manifold reductions”](#), *J. Diff. Eqn.* **141**, 122–149 (1997).
- [65] B. Sandstede, A. Scheel, and C. Wulff, [“Bifurcations and dynamics of spiral waves”](#), *J. Nonlin. Sci.* **9**, 439–478 (1999).
- [66] E. Siminos, [Recurrent Spatio-temporal Structures in Presence of Continuous Symmetries](#), PhD thesis (School of Physics, Georgia Inst. of Technology, Atlanta, 2009).

- [67] E. Siminos and P. Cvitanović, “Continuous symmetry reduction and return maps for high-dimensional flows”, *Physica D* **240**, 187–198 (2011).
- [68] G. I. Sivashinsky, “Nonlinear analysis of hydrodynamical instability in laminar flames - I. Derivation of basic equations”, *Acta Astronaut.* **4**, 1177–1206 (1977).
- [69] S. Smale, “Topology and mechanics, I.” *Inv. Math.* **10**, 305–331 (1970).
- [70] V. Thümmel, *Numerical Analysis of the Method of Freezing Traveling Waves*, PhD thesis (Bielefeld Univ., 2005).
- [71] V. Y. Toronov and V. L. Derbov, “Geometric phases in lasers and liquid flows”, *Phys. Rev. E* **49**, 1392–1399 (1994).
- [72] V. Y. Toronov and V. L. Derbov, “Geometric-phase effects in laser dynamics”, *Phys. Rev. A* **50**, 878–881 (1994).
- [73] A. Vierkandt, “Über Gleitende und Rollende Bewegung”, *Monatsh. Math. Phys.* **III**, 31–54 (1892).
- [74] D. Viswanath, “Recurrent motions within plane Couette turbulence”, *J. Fluid Mech.* **580**, 339–358 (2007).
- [75] A. G. Vladimirov, V. Y. Toronov, and V. L. Derbov, “The complex Lorenz model: Geometric structure, homoclinic bifurcation and one-dimensional map”, *Int. J. Bifur. Chaos* **8**, 723–729 (1998).
- [76] H. Wedin and R. R. Kerswell, “Exact coherent structures in pipe flow”, *J. Fluid Mech.* **508**, 333–371 (2004).
- [77] A. T. Winfree, “Scroll-shaped waves of chemical activity in 3 dimensions”, *Science* **181**, 937–939 (1973).
- [78] A. T. Winfree, *The Geometry of Biological Time* (Springer, New York, 1980).
- [79] J. G. Yoder, *Unrolling Time: Christiaan Huygens and the Mathematization of Nature* (Cambridge Univ. Press, Cambridge UK, 1989).
- [80] A. N. Zaikin and A. M. Zhabotinsky, “Concentration wave propagation in 2-dimensional liquid-phase self-oscillating system”, *Nature* **225**, 535–537 (1970).

13.7 Examples

Example 13.1. An SO(2) moving frame. (Continued from example 12.5) The SO(2) action



$$(\hat{x}_1, \hat{y}_1) = (x_1 \cos \theta + y_1 \sin \theta, -x_1 \sin \theta + y_1 \cos \theta) \quad (13.25)$$

is regular on $\mathbb{R}^2 \setminus \{0\}$. Thus we can define a slice as a ‘hyperplane’ (here a mere line), through $\hat{x}' = (0, 1)$, with group tangent $t' = (1, 0)$, and ensure uniqueness by clockwise rotation into positive y_1 axis. Hence the reduced state space is the half-line $x_1 = 0$, $\hat{x}_2 = y_1 > 0$. The slice condition then simplifies to $\hat{x}_1 = 0$ and yields the explicit formula for the moving frame parameter

$$\theta(x_1, y_1) = \tan^{-1}(x_1/y_1), \quad (13.26)$$

i.e., the angle which rotates the point (x_1, y_1) back to the slice, taking care that $\tan^{-1}$ distinguishes (x_1, y_1) plane quadrants correctly. Substituting (13.26) back to (13.25) and using $\cos(\tan^{-1} a) = (1 + a^2)^{-1/2}$, $\sin(\tan^{-1} a) = a(1 + a^2)^{-1/2}$ confirms $\hat{x}_1 = 0$. It also yields an explicit expression for the transformation to variables on the slice,

$$\hat{x}_2 = \sqrt{x_1^2 + y_1^2}. \quad (13.27)$$

This was to be expected as SO(2) preserves lengths, $x_1^2 + y_1^2 = \hat{x}_1^2 + \hat{y}_1^2$. If dynamics is in plane and SO(2) equivariant, the solutions can only be circles of radius $(x_1^2 + y_1^2)^{1/2}$, so this is the “rectification” of the harmonic oscillator by a change to polar coordinates, example A2.1. Still, it illustrates the sense in which the method of moving frames yields group invariants. (E. Siminos)

[return to p. 238](#)

Example 13.2. In-slice stability. So far we have managed to formulate a relatively simple symmetry reduction method, applicable to many problems of interest. Can we also compute the linear stability of a relative equilibrium within a slice? The answer is yes. We compute the reduced stability matrix by taking the in-slice partial derivative of the in-slice velocity (13.7),



$$\begin{aligned} \frac{\partial \hat{v}(\hat{x})_i}{\partial \hat{x}_j} &= \frac{\partial}{\partial \hat{x}_j} \left\{ v(\hat{x})_i - \frac{\langle v(\hat{x})|t' \rangle}{\langle t(\hat{x})|t' \rangle} t(\hat{x})_i \right\} \\ \hat{A}(\hat{x})_{ij} &= A(\hat{x})_{ij} - \frac{t(\hat{x})_i \{ \langle \langle t(\hat{x})|t' \rangle A(\hat{x})^T - \langle v(\hat{x})|t' \rangle \mathbf{T}^T \}_j}{\langle t(\hat{x})|t' \rangle^2} - \frac{\langle v(\hat{x})|t' \rangle}{\langle t(\hat{x})|t' \rangle} \mathbf{T}_{ij}, \end{aligned} \quad (13.28)$$

or, in matrix notation,

$$\hat{A}(\hat{x}) = A(\hat{x}) - \frac{|t(\hat{x})\rangle \langle \langle t(\hat{x})|t' \rangle A(\hat{x})^T - \langle v(\hat{x})|t' \rangle \mathbf{T}^T \rangle t'|}{\langle t(\hat{x})|t' \rangle^2} - \frac{\langle v(\hat{x})|t' \rangle}{\langle t(\hat{x})|t' \rangle} \mathbf{T}. \quad (13.29)$$

How come we got this lengthy formula (13.29), while the stability of a relative equilibrium looked beautifully simple in (12.28)? That formula was written in the co-rotating frame of a relative equilibrium, i.e., in the language of slicing, we picked the relative equilibrium as the slice template. Substitute $\hat{x} = \hat{x}' = x_{TW}$ in (13.29) and recall that $\langle v(x_{TW})|t(x_{TW}) \rangle = c$. Hence the second term vanishes, and for the relative equilibrium we recover (12.28):

$$\hat{A}(x_{TW}) = A(x_{TW}) - c\mathbf{T}. \quad (13.30)$$

This equation is only true for a relative equilibrium in its own slice. In a general slice, such as the one we described in sect. 13.5, one has to use (13.29) to compute the reduced stability matrix.

(N.B. Budanur)

[return to p. 244](#)

Exercises

- 13.1. **SO(2) or harmonic oscillator slice:** Construct a moving frame slice for action of SO(2) on $\mathbb{R}^2$

$$(x, y) \mapsto (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta)$$

by, for instance, the positive y axis: $x = 0, y > 0$. Write out explicitly the group transformation that brings any point back to the slice. What invariant is preserved by this construction? (E. Siminos)

- 13.2. **The moving frame flow stays in the reduced state space:** Show that the flow (13.7) stays in a $(d-1)$ -dimensional slice hyperplane.
- 13.3. **Stability of a relative equilibrium in the reduced state space:** Find an expression for the stability matrix of the system at a relative equilibrium when a linear slice is used to reduce the symmetry of the flow.
- 13.4. **Stability of a relative periodic orbit in the reduced state space:** Find an expression for the Jacobian matrix (monodromy matrix) of a relative periodic orbit when a linear slice is used to reduce the dynamics of the flow.
- 13.5. **Determination of invariants by the method of slices:** Show that the $d - N$ reduced state space coordinates determined by the method of slices are independent and invariant under group actions, and that the method of slices allows the determination of (in general non-polynomial) symmetry invariants by a simple algorithm that works well in high-dimensional state spaces.
- 13.6. **Invariant subspace of the two-modes system:** Show that $(0, 0, x_2, y_2)$ is a flow invariant subspace of the two-modes system (12.40), i.e., show that a trajectory with the initial point within this subspace remains within it forever.
(N.B. Budanur)
- 13.7. **Slicing the two-modes system:** Choose the simplest slice template point that fixes the 1. Fourier mode,

$$\hat{x}' = (1, 0, 0, 0). \quad (13.31)$$

- (a) Show for the two-modes system (12.40), that the velocity within the slice (13.7), and the phase velocity (13.8) along the group orbit are

$$\hat{v}(\hat{x}) = v(\hat{x}) - \dot{\phi}(\hat{x})t(\hat{x}) \quad (13.32)$$

$$\dot{\phi}(\hat{x}) = -v_2(\hat{x})/\hat{x}_1 \quad (13.33)$$

- (b) Determine the chart border (the locus of point where the group tangent is either not transverse to the slice or vanishes).
- (c) What is its dimension?
- (d) What is its relation to the invariant subspace of exercise 13.7?
- (e) Can a symmetry-reduced trajectory cross the chart border?

(N.B. Budanur and P. Cvitanović)

- 13.8. **The symmetry reduced two-modes flow:** Pick an initial point $\hat{x}(0)$ that satisfies the slice condition (13.4) for the template choice (13.31) and integrate (13.32) & (13.33). Plot the three dimensional slice hyperplane spanned by (x_1, x_2, y_2) to visualize the symmetry reduced dynamics. Does it look like figure 13.5 (b)?
(N.B. Budanur)

- 13.9. **Visualize the relative equilibrium of the two-modes system:** Starting the initial condition

$$x_0 = (0.439966, 0, -0.386267, 0.070204) \quad (13.34)$$

integrate the full state space SO(2)-equivariant (12.40) and the symmetry reduced (13.32) two-modes system for $t = 250$ time units. Plot the (x_1, x_2, y_1) projection of both trajectories. Explain your results.

(N.B. Budanur)

- 13.10. **Relative equilibria of the two-modes system:** Write down an expression for the reduced velocity (13.32) of the two-modes system explicitly by substituting (13.33) and solve $\hat{v} = 0$ to find the relative equilibria. Part of this might be doable analytically (you have an invariant subspace). If that does not work out for you, solve the system numerically, for the parameter values (12.39). Check that x_0 of exercise 13.34 is among your solutions. Mark the relative equilibria that you have found on the strange attractor plot of exercise 13.8, interpret the role they play in the dynamics, if any.
(N.B. Budanur)

- 13.11. **Stability of the two-modes relative equilibrium:**

- (a) Write down the stability matrix of the two-modes system in the reduced state space by computing derivatives of (13.32).
- (b) Compute eigenvalues and eigenvectors of this stability matrix at the relative equilibrium (13.34)
- (c) Indicate the direction along which the nearby trajectories expand.
- (d) Compute the stability eigenvalues and eigenvectors of all relative equilibria of exercise 13.10

(N.B. Budanur)

- 13.12. **Relative periodic orbits of the two-modes system:** Initial conditions and periods for 4 relative periodic orbit of the two-modes system are listed in the table 12.1. Integrate (12.40) and (13.32) with these initial conditions for 3-4 periods, and plot the four trajectories. Explain what you see.

(N.B. Budanur)

- 13.13. **Poincaré section in the slice** Construct a Poincaré section for the two-modes system in the slice hyperplane, such that the relative equilibrium (13.34) and its expanding direction that you found in (13.11) is in this Poincaré section. Interpolate this Poincaré section with a smooth curve, and compute the arclengths positions of each crossing of the symmetry-reduced flow with the Poincaré section.
(N.B. Budanur)

- 13.14. **Finding relative periodic orbits from a return map.** Produce a return map of the arclengths that you found in exercise 13.13. Plot this return map. Note that its derivative is discontinuous at its critical point - why? Interpolate to this return map in two pieces and find its fixed point. Take the fixed point as the initial point to integrate the reduced two-modes system (13.32) for $t = 3.7$. What do you see?
(N.B. Budanur)

Chapter 14

Charting the state space

The classification of the constituents of a chaos, nothing less is here essayed.

—Herman Melville, *Moby Dick*, chapter 32

IN THIS CHAPTER and the next we learn to *partition* state space in a topologically invariant way, and *identify* topologically distinct orbits.

We start in sect. 14.1 with a simple and intuitive example, a 3-disk game of pinball. The qualitative dynamics of stretching/shrinking strips of surviving state space regions enables us to partition the state space and assign *symbolic dynamics* itineraries to trajectories. For the 3-disk game of pinball all possible symbol sequences enumerate all possible orbits. 

In sect. 14.2 we use Rössler and Lorenz flows to motivate modeling of higher-dimensional flows by iteration of 1-dimensional maps. For these two flows the 1-dimensional maps capture essentially all of the higher-dimensional flow dynamics, both qualitatively and quantitatively. 1-dimensional maps suffice to explain the two key aspects of qualitative dynamics; *temporal ordering*, or *itinerary* with which a trajectory visits state space regions (sect. 14.3), and the *spatial ordering* between trajectory points (sect. 14.4), which is the key to determining the admissibility of an orbit with a prescribed itinerary. In a generic dynamical system not every symbol sequence is realized as a dynamical trajectory; as one looks further and further, one discovers more and more ‘pruning’ rules which prohibit families of itineraries. For 1-dimensional ‘stretch & fold’ maps the *kneading theory* (sect. 14.5) provides the definitive answer as to which temporal itineraries are *admissible* as trajectories of the dynamical system. Finally, sect. 14.6 is meant to serve as a guide to the basic concepts of symbolic dynamics. 

Deceptively simple, this subject can get very difficult very quickly, so in this chapter we do the first, 1-dimensional pass at a pedestrian level, postponing the discussion of higher-dimensional, cyclist level issues to chapter 15.

Even if you only care about the serious stuff (like Rydberg atoms or meso-

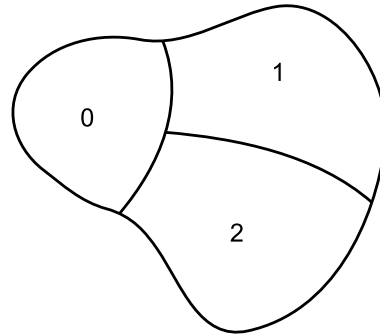


Figure 14.1: A coarse partition of $\mathcal{M}$ into regions $\mathcal{M}_0$, $\mathcal{M}_1$, and $\mathcal{M}_2$, labeled by ternary alphabet $\mathcal{A} = \{0, 1, 2\}$.

scopic devices) and you resent wasting time on formal things, this chapter and chapters 17 and 18 are good for you. Study them.

14.1 Qualitative dynamics

(R. Mainieri and P. Cvitanović)



What can a flow do to points in state space? This is a very difficult question to answer because we have assumed very little about the evolution function f^t . It's important to know if f^t is continuous and if it can be differentiated a sufficient number of times. Trying to make sense of this question is one of the basic concerns in the study of dynamical systems. The first answer was inspired by the motion of the planets: they appear to repeat their motion through the firmament, so the ancients attempted to describe dynamical systems as periodic.



However, periodicity is almost never quite exact. What one tends to observe is *recurrence*. A recurrence of a point x_0 of a dynamical system is a return of that point to a neighborhood of where it started. How close the point x_0 must return is up to us. We can choose a volume of any size and shape, and call it the neighborhood $\mathcal{M}_0$, as long as it encloses x_0 . For chaotic dynamical systems, the evolution might bring the point back to the starting neighborhood infinitely often. That is, the set

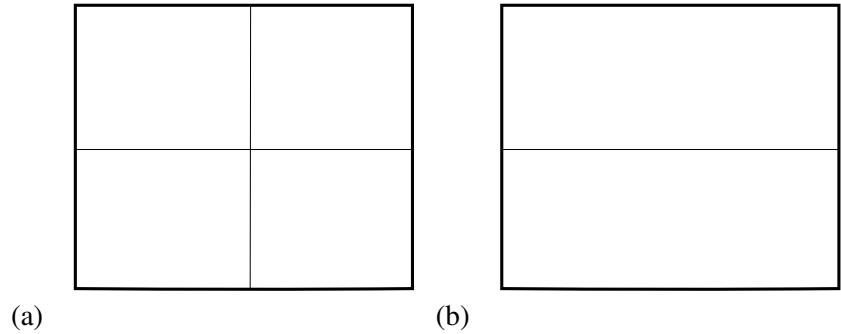
$$\{y \in \mathcal{M}_0 : y = f^t(x_0), \quad t > t_0\} \quad (14.1)$$

will in general have an infinity of recurrent episodes.

To observe a recurrence we must look at neighborhoods of points. This suggests another way of describing how points move in state space, which is an important first step on the way to developing a theory of dynamical systems: qualitative, topological dynamics, or *symbolic dynamics*. As the subject can get quite technical, a summary of the basic notions and definitions of symbolic dynamics is relegated to sect. 14.6; check that section and references cited in remark 14.1, whenever you run into baffling jargon.

We start by dividing the state space into regions $\mathcal{M}_A, \mathcal{M}_B, \dots, \mathcal{M}_Z$, as in figure 14.1. This can be done in many ways, not all equally clever. Any such division

Figure 14.5: The 3-disk game of pinball Poincaré section; trajectories emanating from disk 1 with $x = (\text{arclength, parallel momentum}) = (s, p)$, where $p = \sin \theta$. (a) Strips of initial points $\mathcal{M}_{12}$, $\mathcal{M}_{13}$ which reach disks 2, 3 in one bounce, respectively. (b) 1-step memory refinement of partition (see figure 14.3): strips of initial points $\mathcal{M}_{121}$, $\mathcal{M}_{131}$, $\mathcal{M}_{132}$ and $\mathcal{M}_{123}$ which reach disks 1, 2, 3 in two bounces, respectively. Disk radius : center separation ratio $a:R = 1:2.5$. (Y. Lan)



If there is no way to reach partition $\mathcal{M}_i$ from partition $\mathcal{M}_j$, and conversely, partition $\mathcal{M}_j$ from partition $\mathcal{M}_i$, the state space consists of at least two disconnected pieces, and we can analyze it piece by piece. An interesting partition should be dynamically connected, i.e., one should be able to go from any region $\mathcal{M}_i$ to any other region $\mathcal{M}_j$ in a finite number of steps. A dynamical system with such a partition is said to be *metrically indecomposable*.

In general, one also encounters transient regions - regions to which the dynamics does not return once the trajectory has exited the given state space. Hence we have to distinguish between (uninteresting to us) wandering trajectories that never return to the initial neighborhood, and the non-wandering set (2.5) of the *recurrent* trajectories. We are implicitly assuming that the transients are sufficiently short-lived not to be of experimental interest.

However, knowing that a point from $\mathcal{M}_i$ reaches $\{\mathcal{M}_j, \dots, \mathcal{M}_k\}$ in one step is not quite good enough. We would be happier if we knew that the map of the entire initial region $f(\mathcal{M}_i)$ overlaps nicely with the entire $\mathcal{M}_j$; otherwise we have to subpartition $\mathcal{M}_j$ into the subset $f(\mathcal{M}_i)$ and the reminder. Often we will find ourselves partitioning *ad infinitum*, a difficult topic to which we shall return in sect. 15.4.

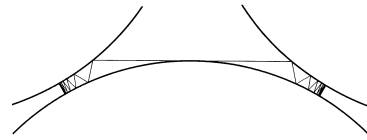
Such considerations motivate the notion of a *Markov partition*, a partition for which no memory of preceding steps is required to fix the transitions allowed in the next step. *Finite Markov partitions* can be generated by *expanding* d -dimensional iterated mappings $f : \mathcal{M} \rightarrow \mathcal{M}$, if $\mathcal{M}$ can be divided into N regions $\{\mathcal{M}_0, \mathcal{M}_1, \dots, \mathcal{M}_{N-1}\}$, such that in one step a set of points from an initial region $\mathcal{M}_i$ either fully cover a region $\mathcal{M}_j$ or miss it altogether,

$$\text{either } \mathcal{M}_j \cap f(\mathcal{M}_i) = \emptyset \quad \text{or} \quad \mathcal{M}_j \subset f(\mathcal{M}_i). \quad (14.2)$$

Whether such partitions can be found is not clear at all - the borders need to be lower-dimensional sets invariant under dynamics, and there is no guarantee that these are topologically simple objects. However, the game of pinball (and many other non-wandering repeller sets) is especially nice. The issue of determining the partition borders does not arise, as the survivors live on disconnected pieces of the state space, separated by a chasm of escaping trajectories.

The itinerary of a billiard trajectory is described by one of the following situations: (1) finite for a scattering trajectory, coming in from infinity and escaping

Figure 14.6: For the 3-disk game of pinball no itineraries are pruned as long as the inter-disk spacing exceeds $R : a > 2.04821419\dots$ (from K.T. Hansen [26])



after a finite number of collisions, (2) infinite for a trapped trajectory, or (3) infinitely repeating for a periodic orbit. A finite length trajectory is not uniquely specified by its finite itinerary, but an isolated unstable cycle is. Its itinerary is an infinitely repeating block of symbols. For hyperbolic flows, the intersection of the future and past itineraries, the bi-infinite itinerary $\cdots s_{-2}s_{-1}s_0.s_1s_2s_3\cdots$, specifies a unique orbit. Almost all infinite length trajectories (orbits) are aperiodic. Still, the longer the trajectory, the more closely it resembles a periodic orbit whose itinerary shadows the trajectory for its whole length. Think of the state space as the unit interval, aperiodic orbits as normal numbers, and periodic ones as fractions whose denominators correspond to cycle periods (as is, for example, literally the case for the Farey map, to be discussed in sect. 29.3.4).

Determining whether the symbolic dynamics is complete (as is the case for sufficiently separated disks, see figure 14.6), pruned (for example, touching or overlapping disks), or only a first coarse-graining of the topology (for example, smooth potentials with islands of stability) requires a case-by-case investigation. This is a discussion we postpone until sect. 14.5 and chapter 15. For now, we assume that the disks are sufficiently separated that there is no additional pruning beyond the prohibition of self-bounces.

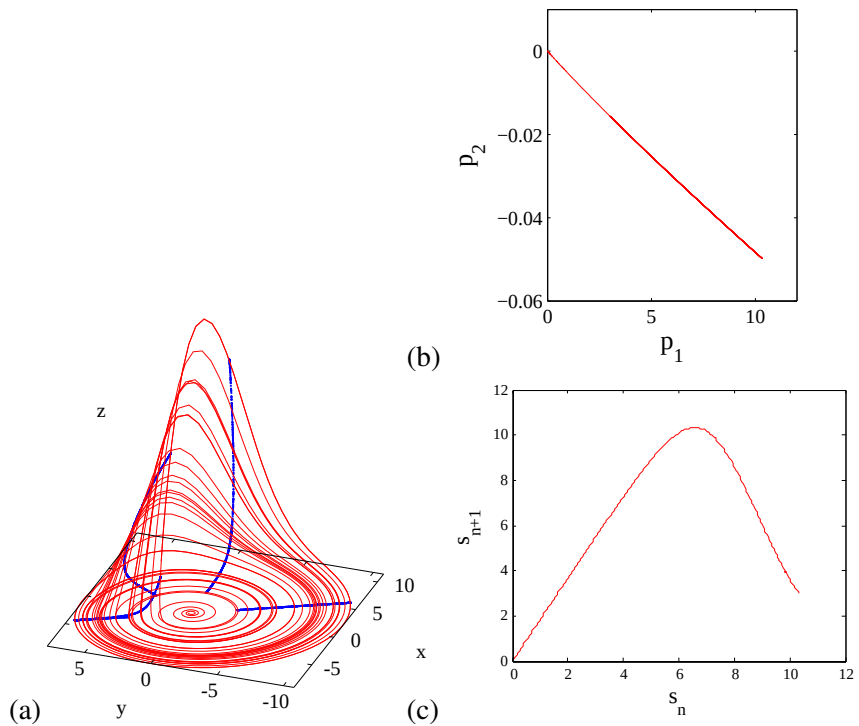
Inspecting figure 14.5, we see that the relative ordering of regions with differing finite itineraries is a qualitative, topological property of the flow. This observation motivates searches for simple, ‘canonical’ partitions which exhibit in a simple manner the spatial ordering common to entire classes of topologically similar nonlinear flows.

14.2 From d -dimensional flows to 1-dimensional maps

Symbolic dynamics for the 3-disk game of pinball is so straightforward that one may altogether fail to see the connection between the topology of hyperbolic flows and their symbolic dynamics. This is brought out more clearly by the 1-dimensional visualization of ‘stretch & fold’ flows to which we turn now.

We construct here the return maps (3.4) for two iconic flows, the Rössler and the Lorenz, in order to show how ODEs in higher dimensions can be modeled by low-dimensional maps. In these examples, the strong dissipation happens to render the dynamics essentially 1-dimensional, both qualitatively and quantitatively. However, as we shall show in chapter 15, strong dissipation is not essential, but hyperbolicity is essential. Thus, this method applies to Hamiltonian (symplectic areas preserving) flows as well. The key idea is to replace the original, arbitrarily concocted coordinates by intrinsic, dynamically invariant curvilinear coordinates erected on neighborhoods of unstable manifolds.

Figure 14.7: (a) The Rössler flow, figure 3.3, is an example of a recurrent flow that stretches and folds. Shift the origin to equilibrium x_- computed in (2.31), $(x, y, z) = (p_0 - x_-, p_1 - y_-, p_2 - z_-)$. (b) $p_0 = 0, p_1 > 0$ Poincaré section of the x_- unstable manifold. (c) $s \rightarrow P(s)$ Rössler ‘stretch & fold’ return map, where s is the arc-length distance measured along the Poincaré section of unstable manifold of equilibrium point x_- . See also figure 14.12. (R. Paškauskas, A. Basu and J. Newman)



fast track:
sect. 14.3, p. 266

Suppose concentrations of certain chemical reactants worry you, or the variations in the Vladivostok temperature, humidity, pressure and winds affect your mood. Such quantities vary within some fixed range, and so do their rates of change. Even if we are studying an open system such as the 3-disk pinball game, we tend to be interested in a finite region around the disks and ignore the escapees. So a typical dynamical system that we care about is *bounded*. If the price to keep going is high, the dynamics tend to settle into a simple state. As an analogy, if we stir up some tar, we observe the dynamics come to a dead stop the moment we cease our labors. However, as the resistance to change decreases with vigorous stirring, the tar heats up, and the dynamics becomes unstable. What happens next?



Just by looking at figure 14.7, you get the idea - Rössler flow winds around the stable manifold of the ‘central’ equilibrium, stretches and folds, and the dynamics on the Poincaré section of the flow can be reduced to a 1-dimensional map.



example 14.3
p. 282



The next, Lorenz flow example is similar, but the folding mechanism is very different. Here the unstable manifold of one equilibria collides with the stable manifold of another one, forcing a robust *heteroclinic connection* between the two.



example 14.4
p. 282

Heteroclinic connections. The simplest example of intersection of invariant manifolds is an orbit on the unstable manifold of an unstable equilibrium that falls into a stable equilibrium (a sink). In general, two manifolds can intersect in a stable way (i.e., robustly with respect to small changes of system parameters), if the sum of their dimensions is greater than or equal to the dimension of the state space; hence, an unstable manifold of dimension k is likely to intersect a stable manifold whose codimension in state space is less than or equal to k . Whether the two manifolds actually intersect is a subtle question that is central to the issue of “structural stability” of ergodic dynamical systems. Trajectories that leave an equilibrium or periodic orbit along its unstable manifold and reach another equilibrium or periodic orbit along its stable manifold are called either *heteroclinic*, if the two invariant solutions are distinct, or *homoclinic*, if the initial and the final invariant solutions are the same.

remark 14.4

What have we learned from the above two exemplary 3-dimensional flows? If a flow is locally unstable but globally bounded, any open ball of initial points will be stretched out and then folded back. If the equilibria are hyperbolic, the trajectories will be attracted along some eigen-directions and ejected along others. The unstable manifold of one equilibrium can avoid stable manifolds of other equilibria, as is the case for Rössler, or plow into them head on, as is the case for Lorenz. A typical trajectory wanders through state space, always attracted to the next equilibrium or periodic orbit neighborhood, and then ejected again. What is important is the motion along the unstable manifolds – that is where 1d maps come from.

At this juncture we proceed to show how this works on the simplest example, unimodal mappings of the interval. The erudite reader may skim through this chapter and then take a more demanding path, via the Smale horseshoes of chapter 15. Unimodal maps are easier, but less physically compelling. Smale horseshoes offer the high road. They are more complicated, but the right tool to generalize the 3-disk dynamics and begin analysis of general dynamical systems. It is up to you - unimodal maps suffice to get quickly to the heart of this treatise.

14.3 Temporal ordering: Itineraries

In this section we learn to *name* topologically distinct trajectories for the simple, but instructive case of the 1-dimensional maps of an interval. The simplest such map is the “coin flip” of figure 14.8; the unit interval is stretched, cut, and overlaid over itself.



example 14.5
p. 284

A more physically motivated mapping of this type is *unimodal*. The interval is stretched and folded only once, with at most two points mapping into a point in the refolded interval, as in the Rössler return map figure 14.7 (b). A unimodal map $f(x)$ is a 1-dimensional function $\mathbb{R} \rightarrow \mathbb{R}$ defined on an interval $M \in \mathbb{R}$ with

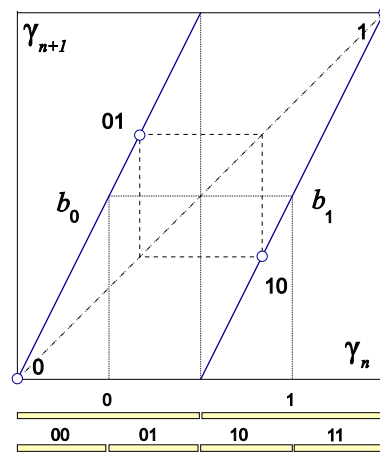
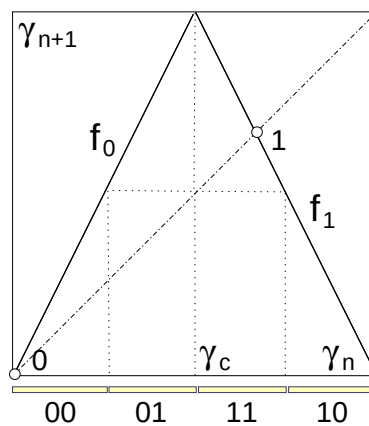


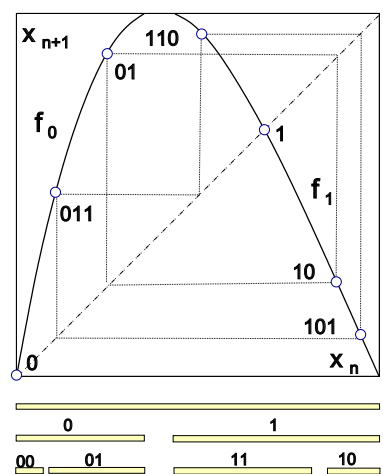
Figure 14.8: The $n = 2$ and 4 intervals state space partitions for the Bernoulli shift map (14.18), together with the fixed points $\bar{0}$, $\bar{1}$ and the 2-cycle $\bar{01}$.

Figure 14.9: (a) The full tent map (14.21) partition $\{\mathcal{M}_{00}, \mathcal{M}_{01}, \mathcal{M}_{11}, \mathcal{M}_{10}\}$ together with the fixed points x_0, x_1 .

(b) A unimodal repeller with the survivor intervals after 1 and 2 iterations. Intervals marked $s_1 s_2 \dots s_n$ consist of points that do not escape in n iterations, and follow the itinerary $S^+ = s_1 s_2 \dots s_n$. Indicated are the fixed points $\bar{0}$, $\bar{1}$, the 2-cycle $\bar{01}$, and the 3-cycle $\bar{011}$. Note that here, unlike the Bernoulli map example of figure 14.8, the spatial ordering does not respect the binary ordering; for example $x_{00} < x_{01} < x_{11} < x_{10}$.



(a)



(b)

a monotonically increasing (or decreasing) branch, a *critical point* (or interval) x_c for which $f(x_c)$ attains the maximum (minimum) value, and followed by a monotonically decreasing (increasing) branch. *Uni*-modal means that the map is a 1-humped map with one critical point within interval $\mathcal{M}$. *Multi*-modal maps, with several critical points within interval $\mathcal{M}$, can be described with a straightforward generalization of the methods we describe next.



example 14.6
p. 284

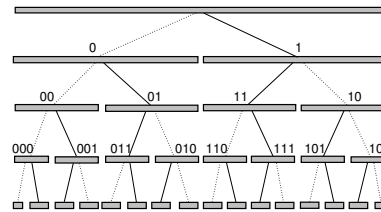


example 14.7
p. 285

For $1d$ maps the *critical value* denotes either the maximum or the minimum value of $f(x)$ on the defining interval; we assume here that it is a maximum, $f(x_c) \geq f(x)$ for all $x \in \mathcal{M}$. The critical point x_c that yields the critical value $f(x_c)$ belongs to neither the left nor the right partition $\mathcal{M}_i$ and is instead denoted by its own symbol $s = C$. As we shall see, its images and preimages serve as partition boundary points.

The trajectory $x_1, x_2, x_3, \dots$ of the initial point x_0 is given by the iteration $x_{n+1} = f(x_n)$. Iterating f and checking whether the point lands to the left or to the right of x_c generates a *temporally* ordered topological itinerary (14.9) for a given

Figure 14.10: An alternating binary tree relates the itinerary labeling of the unimodal map intervals, figure 14.9, to their spatial ordering. The dotted line stands for 0, the full line for 1; the binary sub-tree whose root is a full line with symbol 1 reverses the orientation, due to the orientation-reversing fold in figures 14.7 and 14.9. See also figure 17.5.



trajectory,

$$s_n = \begin{cases} 1 & \text{if } x_n > x_c \\ C & \text{if } x_n = x_c \\ 0 & \text{if } x_n < x_c \end{cases} . \quad (14.3)$$

We refer to $S^+(x_0) = s_1 s_2 s_3 \cdots$ as the *future itinerary*. Our next task is to answer the reverse problem: given an itinerary, what is the *spatial* ordering of points that belong to the corresponding state space trajectory?

14.4 Spatial ordering

A well-known theorem states that combinatorial factors are impossible to explain [1].

—G. 't Hooft and M. Veltman, DIAGRAMMAR

Suppose you have succeeded in constructing a covering symbolic dynamics, such as the one we constructed for a well-separated 3-disk system. Now start moving the disks toward each other. At some critical separation (see figure 14.6), a disk will start blocking families of trajectories traversing the other two disks. The order in which trajectories disappear is determined by their relative ordering in space; the ones closest to the intervening disk will be pruned first. Determining inadmissible itineraries requires that we relate the spatial ordering of trajectories to their time ordered itineraries.

exercise 15.8

The easiest point of departure is to start by working out this relation for the symbolic dynamics of 1-dimensional mappings. As it appears impossible to present this material without getting bogged down in a sea of 0's, 1's and subscripted subscripts, we announce the main result before embarking upon its derivation:

The admissibility criterion (sect. 14.5) eliminates *all* itineraries that cannot occur for a given unimodal map.

section 14.5

For the Bernoulli shift, converting itineraries into a topological ordering is easy; the binary expansion of coordinate γ is also its temporary itinerary. For the tent map (see equation (14.21) and figure 14.9 (a)), it is a bit harder. This map consists of two straight segments joined at $x = 1/2$. The symbol s_n defined in (14.3) equals

0 if the function increases, and 1 if it decreases. Iteration forward in time generates the time itinerary. More importantly, the piecewise linearity of the map makes the converse possible. Thus, one can analytically determine an initial point given its itinerary, a property that we now use to define a topological coordinatization common to all unimodal maps.



Here we have to face the fundamental problem of pedagogy; combinatorics cannot be taught. The best one can do is state the answer and hope that you will figure it out by yourself.



The tent map point $\gamma(S^+)$ with future itinerary S^+ is given by converting the itinerary of s_n 's into a binary number γ by the following algorithm:

$$w_{n+1} = \begin{cases} w_n & \text{if } s_{n+1} = 0 \\ 1 - w_n & \text{if } s_{n+1} = 1 \end{cases}, \quad w_1 = s_1$$

$$\gamma(S^+) = 0.w_1w_2w_3\dots = \sum_{n=1}^{\infty} w_n/2^n. \quad (14.4)$$

This follows by inspection from the binary tree of figure 14.10. Once you figure this out, feel free to complain that the way the rule is stated here is incomprehensible, and show us how you did it better.

exercise 14.4



example 14.9
p. 286

We refer to $\gamma(S^+)$ as the (*future*) *topological coordinate*. The w_i 's are the digits in the binary expansion of the starting point γ for the full tent map in figure 14.9 (a) (see (14.21)). In the left half-interval, the map $f(x)$ acts as multiplication by 2. Likewise, in the right half-interval, the map acts as a flip as well as multiplication by 2, reversing the ordering, and generating in the process the sequence of s_n 's from the binary digits w_n .

The mapping $x_0 \rightarrow S^+(x_0) \rightarrow \gamma_0 = \gamma(S^+)$ is a *topological conjugacy*. Here, the trajectory of the initial point x_0 is mapped to the trajectory of the initial point γ_0 , by iteration of the given unimodal map. Each trajectory has the same itinerary as the 'canonical' unimodal map, i.e., the full tent map (14.21). The virtue of this conjugacy is that $\gamma(S^+)$ *preserves the ordering* for any unimodal map in the sense that if $y > x$, then $\gamma(S^+(y)) > \gamma(S^+(x))$.



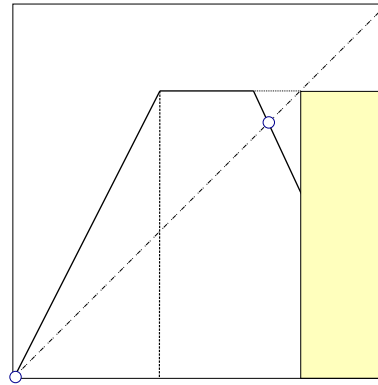
example 14.8
p. 285



example 14.10
p. 286

Critical points are special - they define the boundary between intervals, i.e., the state space is split into M_0 [left part], x_c [critical point] and M_1 [right part] intervals. For the dike map figure 14.11 and the repeller figure 14.9, x_c is the whole interval of points along the flat top of the map, but usually it is a point. As illustrated by figures 14.9 and 14.8, for a unimodal map the preimages $f^{-n}(x_c)$ of the critical point x_c serve as partition boundary points. But this is not true for all preimages—one has to ensure that they are within the set of all admissible orbits

Figure 14.11: The dike map is obtained by slicing off the top portion of the tent map in figure 14.9(a). Any full tent map orbit that visits the primary pruning interval $(\kappa, 1]$ is inadmissible. The admissible orbits form the Cantor set obtained by removing from the unit interval the primary pruning interval and all its iterates. Any admissible orbit has the same topological coordinate and itinerary as the corresponding full tent map orbit.



by checking them against the kneading sequence of the map, to be explained next.

14.5 Kneading theory

No, you can't always get what you want
 You can't always get what you want
 You can't always get what you want
 But if you try sometime you find
 You get what you kneed
 —Bradford Taylor

(K.T. Hansen and P. Cvitanović)

The reason we need to be mindful of spatial ordering of temporal itineraries is that the spatial ordering provides us with criteria that separate inadmissible orbits from those realizable by the dynamics. For 1-dimensional mappings the *kneading theory* provides a precise and definitive criterion of admissibility.



If the parameter in the quadratic map (14.20) is $A > 4$, or the top of unimodal map in figure 14.9 exceeds 1, then the iterates of the critical point x_c diverge for $n \rightarrow \infty$, and any sequence S^+ composed of letters $s_i = \{0, 1\}$ is admissible, and any value of $0 \leq \gamma < 1$ corresponds to an admissible orbit in the non-wandering set of the map. The corresponding repeller is a complete binary labeled Cantor set, the $n \rightarrow \infty$ limit of the n th level covering intervals sketched in figure 14.9.



For $A < 4$ only a subset of the points in the interval $\gamma \in [0, 1]$ corresponds to admissible orbits. The forbidden symbolic values are determined by observing that the largest x_n value in an orbit $x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow \dots$ has to be smaller than or equal to the image of the critical point, the *critical value* $f(x_c)$. Let $K = S^+(x_c)$ be the itinerary of the critical point x_c , denoted by the *kneading sequence* of the map. The corresponding topological coordinate is called the *kneading value*

$$\kappa = \gamma(K) = \gamma(S^+(x_c)). \quad (14.5)$$

S	$\hat{\gamma}(S)$	S	$\hat{\gamma}(S)$
$\overline{0}$	$\overline{.0} = 0$	$\overline{10111}$	$\overline{.11010} = 26/31$
$\overline{1}$	$\overline{.10} = 2/3$	$\overline{10110}$	$\overline{.1101100100} = 28/33$
$\overline{10}$	$\overline{.1100} = 4/5$	$\overline{10010}$	$\overline{.11100} = 28/31$
$\overline{101}$	$\overline{.1110} = 6/7$	$\overline{10011}$	$\overline{.1110100010} = 10/11$
$\overline{100}$	$\overline{.111000} = 8/9$	$\overline{10001}$	$\overline{.11110} = 30/31$
$\overline{1011}$	$\overline{.11010010} = 14/17$	$\overline{10000}$	$\overline{.1111100000} = 32/33$
$\overline{1001}$	$\overline{.11110} = 14/15$		
$\overline{1000}$	$\overline{.11110000} = 16/17$		

Table 14.1: The maximal values of unimodal map cycles up to period 5. (K.T. Hansen)

The ‘canonical’ map that has the same kneading sequence K as $f(x)$ is the dike map, figure 14.11,

$$f(\gamma) = \begin{cases} f_0(\gamma) = 2\gamma & \gamma \in \mathcal{M}_0 = [0, \kappa/2) \\ f_c(\gamma) = \kappa & \gamma \in \mathcal{M}_c = [\kappa/2, 1 - \kappa/2] \\ f_1(\gamma) = 2(1 - \gamma) & \gamma \in \mathcal{M}_1 = (1 - \kappa/2, 1] \end{cases}, \quad (14.6)$$

obtained by slicing off all $\gamma(S^+(x_0)) > \kappa$. The dike map is the full tent map figure 14.9(a) with the top sliced off. It is convenient for coding the symbolic dynamics, as those γ values that survive the pruning are the same as for the full tent map figure 14.9(a), and are easily converted into admissible itineraries by (14.4).

If $\gamma(S^+) > \gamma(K)$, the point x whose itinerary is S^+ would exceed the critical value, $x > f(x_c)$, and hence cannot be an admissible orbit. Let

$$\hat{\gamma}(S^+) = \sup_m \gamma(\sigma^m(S^+)) \quad (14.7)$$

be the *maximal value*, the highest topological coordinate reached by the orbit $x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow \dots$, where σ is the shift (see (14.12)), $\sigma(.s_1s_2s_3\cdots) = .s_2s_3\cdots$. For cycles up to period 5, the maximal values are listed in table 14.1. We shall call the interval $(\kappa, 1]$ the *primary pruned interval*. The orbit S^+ is inadmissible if γ of any shifted sequence of S^+ falls into this interval. 

Criterion of admissibility: Let κ be the kneading value of the critical point, and $\hat{\gamma}(S^+)$ be the maximal value of the orbit S^+ . Then the orbit S^+ is admissible if and only if $\hat{\gamma}(S^+) \leq \kappa$.

question 14.1

While a particular unimodal map may depend on many parameters, its dynamics determines the unique kneading value κ . We shall call κ the *topological parameter* of the map. Unlike the parameters of the original dynamical system, the topological parameter has no reason to be either smooth or continuous. The jumps in κ as a function of the map parameter, such as A in (14.20), correspond to inadmissible values of the topological parameter. Each jump in κ corresponds to a stability window associated with a stable cycle of a smooth unimodal map. For the quadratic map (14.20), κ increases monotonically with the parameter A , but for a general unimodal map such monotonicity need not hold.

Figure 14.12: (a) Web diagram generated by kneading sequence $K = S^+(x_c)$ (the trajectory of the critical point) for the unimodal Rössler return map of figure 14.7 (c). (b) Return map for the $p_0 = 0, p_1 < 0$ Poincaré section of the x_- unstable manifold. The kneading sequence is the same, as this map is conjugate to figure 14.7 (b) by 180° turn. The section, however, is in the region of strong folding, and the map is less convenient in practice. (A. Basu and J. Newman)

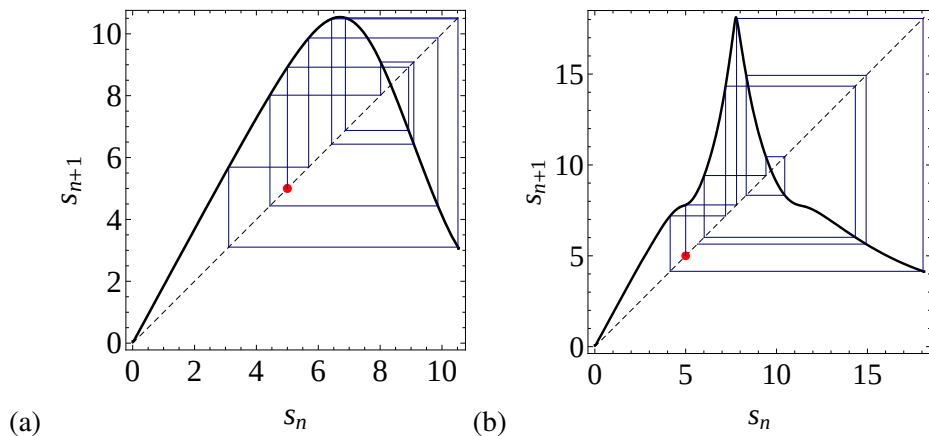
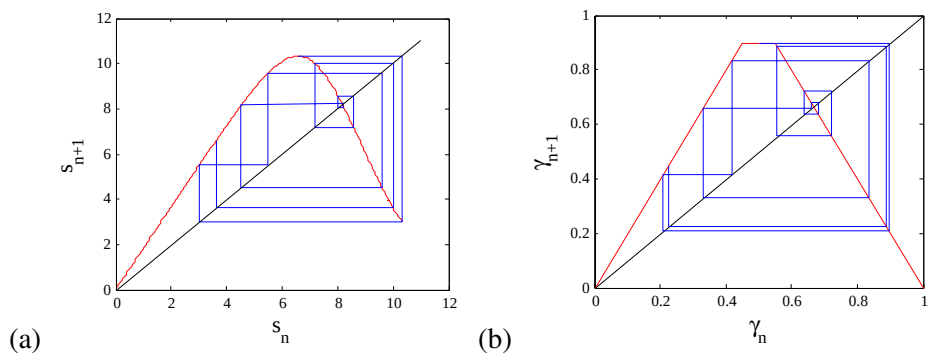


Figure 14.13: (a) Web diagram generated by the trajectory of the critical point, the unimodal Rössler return map of figure 14.7 (b). (b) The web diagram for the corresponding ‘canonical’ dike map (14.6) with the same kneading sequence. (A. Basu and J. Newman)



example 14.11
p. 286

For further details of unimodal dynamics, the reader is referred to appendix A18.1. As we shall see in sect. 15.4, for higher dimensional maps and flows there is no single parameter that orders dynamics monotonically; as a matter of fact, there is an infinity of parameters that need adjustment for a given symbolic dynamics. This difficult subject is beyond our current ambition horizon.



fast track:
chapter 15, p. 290

14.6 Symbolic dynamics, basic notions

(Mathematics) is considered a specialized dialect of the natural language and its functioning as a special case of speech.

— Yuri I. Manin [34]

In this section we collect the basic notions and definitions of symbolic dynamics. The reader might prefer to skim through this material on a first reading and return to it later, as the need arises.



Shifts. We associate with every initial point $x_0 \in \mathcal{M}$ the *future itinerary*, a sequence of symbols $S^+(x_0) = s_1 s_2 s_3 \cdots$ which indicates the order in which the regions are visited. If the trajectory $x_1, x_2, x_3, \dots$ of the initial point x_0 is generated by

$$x_{n+1} = f(x_n), \quad (14.8)$$

then the itinerary is given by the symbol sequence

$$s_n = s \quad \text{if} \quad x_n \in \mathcal{M}_s. \quad (14.9)$$

Similarly, the *past itinerary* $S^-(x_0) = \cdots s_{-2} s_{-1} s_0$ describes the history of x_0 , the order in which the regions were visited before arriving at the point x_0 . For each point x_0 in the state space we thus associate a bi-infinite itinerary

$$S(x_0) = (s_k)_{k \in \mathbb{Z}} = S^- \cdot S^+ = \cdots s_{-2} s_{-1} s_0 \cdot s_1 s_2 s_3 \cdots, \quad (14.10)$$

or simply itinerary, if we chose not to use the decimal point to indicate the present. The itinerary will be finite for a scattering trajectory, entering and then escaping $\mathcal{M}$ after a finite time, infinite for a trapped trajectory, and infinitely repeating for a periodic trajectory.

The set of all bi-infinite itineraries that can be formed from the letters of the alphabet $\mathcal{A}$ is called the *full shift* (or *topological Markov chain*)

$$\mathcal{A}^{\mathbb{Z}} = \{(s_k)_{k \in \mathbb{Z}} : s_k \in \mathcal{A} \text{ for all } k \in \mathbb{Z}\}. \quad (14.11)$$

The jargon is not thrilling, but this is how professional dynamicists talk to each other. We will stick to plain English to the extent possible.

Here we refer to this set of all conceivable itineraries as the *covering* symbolic dynamics. The name *shift* is descriptive of the way the dynamics acts on these sequences. As is clear from the definition (14.9), a forward iteration $x \rightarrow x' = f(x)$ shifts the entire itinerary to the left through the ‘decimal point.’ This operation, denoted by the shift operator σ ,

$$\sigma(\cdots s_{-2} s_{-1} s_0 \cdot s_1 s_2 s_3 \cdots) = \cdots s_{-2} s_{-1} s_0 s_1 \cdot s_2 s_3 \cdots, \quad (14.12)$$

demotes the current partition label s_1 from the future S^+ to the ‘has been’ itinerary S^- . The inverse shift σ^{-1} shifts the entire itinerary one step to the right.

A finite sequence $b = s_k s_{k+1} \cdots s_{k+n_b-1}$ of symbols from $\mathcal{A}$ is called a *block* of length n_b . If the symbols outside of the block remain unspecified, we denote the totality of orbits that share this block by $\_s_k s_{k+1} \cdots s_{k+n_b-1}\_$.

A state space point is a *periodic point* if its orbit returns to it after a finite time; in shift space the orbit is periodic if its itinerary is an infinitely repeating block p^∞ .

We shall refer to the set of periodic points $\mathcal{M}_p$ that belong to a given periodic orbit as a *cycle*

$$p = \overline{s_1 s_2 \cdots s_{n_p}} = \{x_{s_1 s_2 \cdots s_{n_p}}, x_{s_2 \cdots s_{n_p} s_1}, \cdots, x_{s_{n_p} s_1 \cdots s_{n_p-1}}\}. \quad (14.13)$$

A *prime* cycle p of period n_p is a single traversal of the orbit. Its label is a block of n_p symbols that cannot be written as a repeat of a shorter block (in the literature, such cycles are sometimes called *primitive*; we shall refer to it as ‘prime’ throughout this text). By its definition, a cycle is invariant under cyclic permutations of the symbols in the repeating block. A bar over a finite block of symbols denotes a periodic itinerary with an infinitely repeating basic block. We shall omit the bar whenever it is clear from the context that the orbit is periodic. Each *periodic point* is labeled by the starting symbol $s_0 = s_{n_p}$, the next $(n_p - 1)$ steps of its future itinerary. For example, the 2nd periodic point is labeled by

$$\overline{x_{s_1 s_2 \dots s_{n_p}}} = \overline{x_{s_1 s_2 \dots s_0 \cdot s_1 s_2 \dots s_{n_p}}}.$$

This notation (a bit strained) is meant to indicate that the symbol block repeats both in the past and in the future. It is helpful for determining spatial ordering of cycles of 2D-hyperbolic maps, to be undertaken in sect. 15.3.1.

An orbit that starts out as a finite block followed by infinite number of repeats of another block $p = (s_1 s_2 s_3 \dots s_n)$ is said to be *heteroclinic* to the cycle p . An orbit that starts out as p^∞ followed by a different finite block followed by $(p')^\infty$ of another block p' is said to be a *heteroclinic connection* from cycle p to cycle p' . If the orbit returns to the initial cycle (or equilibrium point), $p = p'$, the orbit is said to be a *homoclinic connection*.

Partitions. A partition is called *generating*, if every infinite symbol sequence corresponds to a unique point in state space. Coding is *non-singular* if $x \neq y$, which implies that codes $C(x) \neq C(y)$. The finite Markov partition (14.2) is an example. Constructing a generating partition for a given system is a difficult problem. In the examples to follow, we shall concentrate on cases that permit finite partitions, but in practice almost any generating partition of interest is infinite.

While an infinite itinerary corresponds to a unique point in the state space, any finite itinerary $b = s_{t-\ell+1} \dots s_{-1} s_0 s_1 s_2 \dots s_t$ determines a *cylinder set* $\mathcal{M}_b$. The set of all trajectories $x(t) \in \mathcal{M}$, whose itineraries

$$\dots a_{t-\ell-1} a_{t-\ell} \mathbf{s}_{t-\ell+1} \dots \mathbf{s}_{-1} \mathbf{s}_0 \mathbf{s}_1 \mathbf{s}_2 \dots \mathbf{s}_t a_{t+1} a_{t+2} \dots, \quad (14.14)$$

share the same finite b symbol block, have arbitrary $a_i \in \mathcal{A}$ outside it.

A partition too coarse, coarser than, for example, a Markov partition, would assign the same symbol sequence to distinct dynamical trajectories. To avoid that, we often find it convenient to work with partitions finer than strictly necessary. Ideally the dynamics in the refined partition assigns a unique infinite itinerary $\dots s_{-2} s_{-1} s_0 s_1 s_2 s_3 \dots$ to each distinct orbit, but there might exist full shift symbol sequences (14.11) which are not realized as orbits. These sequences are called *inadmissible*, and we say that the symbolic dynamics is *pruned*. The word is suggested by the ‘pruning’ of branches corresponding to forbidden sequences for symbolic dynamics organized hierarchically into a tree structure, as explained in chapter 17.

A mapping $f : \mathcal{M} \rightarrow \mathcal{M}$ together with a partition $\mathcal{A}$ induces *topological dynamics* (Σ, σ) , where the *subshift*

$$\Sigma = \{(s_k)_{k \in \mathbb{Z}}\}, \quad (14.15)$$

is the set of all *admissible* infinite itineraries, and $\sigma : \Sigma \rightarrow \Sigma$ is the shift operator (14.12). The designation ‘subshift’ comes from the fact that $\Sigma \subset \mathcal{A}^{\mathbb{Z}}$ is the subset of the full shift (14.11). The principal task in developing the symbolic dynamics of a dynamical system that occurs in applications will be to determine Σ , the set of all bi-infinite itineraries S that are actually realized by the given dynamical system.

Pruning. If the dynamics is pruned, the alphabet must be supplemented by a *grammar*, a set of pruning rules. After the inadmissible sequences have been pruned, it is often convenient to parse the symbolic strings into words of variable length - this is called *coding*. Suppose that the grammar can be stated as a finite number of pruning rules, each forbidding a block of finite length,

$$\mathcal{G} = \{b_1, b_2, \dots, b_k\}, \quad (14.16)$$

where a *pruned block* b is a sequence of symbols $b = s_1 s_2 \dots s_{n_b}$, $s \in \mathcal{A}$, of finite length n_b . In this case we can always construct a finite Markov partition (14.2) by replacing finite length words of the original partition by letters of a new alphabet. In particular, if the longest forbidden block is of length $M + 1$, we say that the symbolic dynamics is a shift of finite type with M -step memory. In that case we can *recode* the symbolic dynamics in terms of a new alphabet, with each new letter given by an admissible block of at most length M .

A topological dynamical system (Σ, σ) , for which all admissible itineraries are generated by a finite transition matrix (see (17.1))

$$\Sigma = \{(s_k)_{k \in \mathbb{Z}} : T_{s_k s_{k+1}} = 1 \text{ for all } k\}, \quad (14.17)$$

is called a subshift of *finite type*.



example 14.12
p. 286



in depth:
chapter 15, p. 290

Résumé

What you kneed to know.

—Justin Lanier

In chapters 2 to 4, we focused on things local: a representative point, a short-time trajectory, a neighborhood. In this chapter we have made a courageous leap and gone global.

The main lesson is that - if one intends to go thoughtfully about globalization - one should trust the dynamics itself, and let it partition the state space, by means of

its (topologically invariant) unstable manifolds. This works if every equilibrium and periodic orbit is unstable, so one exits its local neighborhood via its unstable manifold. We delineate the segment of the unstable manifold between the fixed point and the point where the nonlinearity of the dynamics folds back onto itself as the primary segment. We also measure the location of nearby state space points by arclengths measured along this (curvilinear) segment. For 1-dimensional maps, the folding point is the critical point and easy to determine. In higher dimensions, the situation is not so clear - we shall discuss that in chapter 15.

Trajectories exit a neighborhood of an equilibrium or periodic point along unstable directions and fall along stable manifolds towards other invariant orbits, until they again are repelled along their unstable manifolds. Such sequences of visitations can be described by *symbolic dynamics*. As we shall show in chapter 17, they are encoded by transition matrices / transition graphs and approximated dynamically by sequences of unstable manifold $\rightarrow$ unstable manifold maps, or, in case of a return to the initial neighborhood, by return maps $s \rightarrow f(s)$.

As the kneading theory of sect. 14.5 illustrates, not all conceivable symbol sequences are actually realized (*admissible*). The identification of all inadmissible or *pruned* sequences is in general not possible. However, the theory to be developed here relies on exhaustive enumeration of all admissible itineraries up to a given topological length; chapters 15 and 18 describe several strategies for accomplishing this for physically realistic goals.



Commentary

Remark 14.1. Symbolic dynamics. For a brief history of symbolic dynamics, from Hadamard in 1898, Morse and Hedlund in 1938 and onward, see notes to chapter 1 of Kitchens monograph [32], a very clear and enjoyable mathematical introduction to topics discussed here. Diacu and Holmes [20] provide an excellent survey of symbolic dynamics applied to celestial mechanics. For a compact survey of symbolic dynamics techniques, consult sects. 3.2 and 8.3 of Robinson [47]; for a gentler introduction, Rodriguez-Gonzalez [49]. The binary labeling of the once-folding map periodic points was introduced by Myrberg [39–43] for 1-dimensional maps, and its utility to 2-dimensional maps has been emphasized in refs. [23, 38]. For 1-dimensional maps, it is now customary to use the R - L notation of Metropolis, Stein and Stein [17, 36], indicating that the point x_n lies either to the left or to the right of the critical point in figure 14.9. The symbolic dynamics of such mappings has been extensively studied by means of the Smale horseshoes, see for example ref. [24]. Using letters rather than numerals in a symbol dynamics alphabet $\mathcal{A}$ probably reflects good taste. We prefer numerals for their computational convenience, as they speed up conversions of itineraries into the topological coordinates (δ, γ) introduced in sect. 15.3.1. The alternating binary ordering of figure 14.10 is related to the Gray codes of computer science [46]. Kitchens [32] convention is $\cdots s_{-2}s_{-1}.s_0s_1s_2s_3\cdots$, with ‘.’ placed differently from our convention (14.10).

Remark 14.2. Bernoulli map. The Bernoulli shift map (14.18) and the doubling map (14.19) are also known as the dyadic transformation, dyadic map, bit shift map, angle

doubling map or sawtooth map (24.21). There are many fine books that discuss it in depth, for example Driebe [21]. See also remark 28.3.

Remark 14.3. Kneading theory. The admissible itineraries are studied, for example, in refs. [19, 24, 36, 50]. We follow here the Milnor-Thurston exposition [37]. They study the topological zeta function for piecewise monotone maps of the interval, and they also show that the topological zeta function can be expressed in terms of a finite dimensional *kneading determinant* for the case of a finite subshift. As the kneading determinant is essentially the topological zeta function of sect. 18.4, we do not discuss it here. Baladi and Ruelle have reworked this theory in a series of papers [7–9]. See also P. Dahlqvist’s appendix A18.1. Knight and Klages refer to the set of iterates of the critical point as the ‘generating orbit’ in their study of deterministic diffusion [33] (for deterministic diffusion, see chapter 24). They say: “The structure of the Markov partitions varies wildly under parameter variation. The method we employ to understand the Markov partitions involves iterating the critical point. The set of iterates of this point form a set of Markov partition points for the map. Hence we call the orbit of the critical point a ‘generating orbit.’ If the generating orbit is finite for a particular value of parameters, we obtain a finite Markov partition. We can then use the finite Markov partition to tell us about the diffusive properties of the map and hence the structure of the diffusion coefficient.”

Question 14.1. Alice wants to know

Q ‘Criterion of admissibility’? What’s the big deal?

A It is amazing - if you know the symbolic itinerary of *one* trajectory, you know *all* admissible itineraries. Later, when we will need to compute periodic orbits, this will enable us to compute them *all* up to a given period, without any guessing and without the danger of missing some, potentially important ones. Savor this moment, this is the last time you know everything - we will never again have a full description of all possible orbits in any problem that we care about. A bit like passage from teenagehood to adulthood.

Remark 14.4. Heteroclinic connections. For sketches of heteroclinic connections in the nonlinear setting, see Abraham and Shaw illustrated classic [2]. Section 5 of ref. [30] makes elegant use of stable manifold co-dimension counts and of invariant subspaces implied by discrete symmetries of the underlying PDE to deduce the existence of a heteroclinic connection. Ref. [31] which defines heteroclinic connections, cycles and networks has lots of references. It focuses on two-dimensional unstable manifolds, discusses discrete symmetries, robust cycles on invariant subspaces, and constructs ‘cross-sections’ that lie within the region of approximate linear flow near equilibria.

Remark 14.5. Cat map state space partition. (Continued from remark 8.3) The “linear code” [13, 44] for temporal evolution of a single cat map was introduced in the beautiful 1987 paper by Percival and Vivaldi. Percival and Vivaldi give a clear discussion of the “two-configuration cat map representation” in Section 4, *Geometry of the linear code*, of their paper [44]. It is well worth reading. Their alphabet is different, as their map acts on the centered unit interval $[-1/2, 1/2)$. However, digesting Percival and Vivaldi [45] will require some pencil and paper. As explained in ref. [25], the deep problem with the Percival-Vivaldi code prescription is that it does not yield a generating partition; the borders (i.e., x_0, x_1 axes) of their unit-square partition $(x_{t-1}, x_t) \in (0, 1] \times (0, 1]$ do not map onto themselves, resulting in the infinity of, to us unknown, grammar rules for inadmissible symbol sequences.

This problem was resolved in 1967 by Adler and Weiss [3, 4, 6] who utilized the stable/unstable manifolds of the fixed point at the origin to cover a unit area torus by

a two-rectangle generating partition; for the Percival-Vivaldi cat map (8.52), such partition is drawn in figure 14.18. Following Bowen [16], one refers to parallelograms in figure 14.18 as ‘rectangles’; for details see Devaney [19] or Robinson [48]. Siemaszko and Wojtkowski [51] refer to such partitions as the ‘Berg partitions’, and Creagh [18] studies their generalization to weakly nonlinear mappings. Symbolic dynamics on this partition is a subshift of finite type, with the 3-letter alphabet (14.26). While Percival and Vivaldi were well aware of Adler-Weiss partitions, they felt that the coding they introduced “is less efficient in requiring more symbols, but it has the advantage of linearity.” Our construction demonstrates that one can have both: an Adler-Weiss generating cat map partition and a linear code. The only difference from the Percival-Vivaldi formulation [44] is that one trades the single unit-square cover of the torus in (8.52) for the dynamically intrinsic, two-rectangles cover of figure 14.18. This trade has a magical effect - now every infinite walk on the transition graph of figure 14.18 (d) corresponds to a unique admissible orbit $\{x_t\}$, and the transition graph generates all admissible itineraries $\{s_t\}$.

Cat maps admit simple Markov partitions [6, 19], which lead to a simple rational polynomial expression (18.44) for topological zeta functions due to Isola [27], where the numerator corrects the overcounting of the fixed point at the origin due to assigning it to both generating partition rectangles [35]. An extensive analysis of the periodic orbits of the cat map, following Percival and Vivaldi [44], has been carried out by Keating [28]. The fact that even Dyson [22] counts cat map periods should give you pause - clearly, some nontrivial number theory is afoot. Problems with the discretization of Arnol’d cat map were pointed out in refs. [14, 15]. Ref. [15] discusses two partitions of the cat map unit square. There are many good expositions - in particular we enjoyed Robinson [48] and Anosov, Klimenko and Kolutsky [5]. Behrends [11, 12] *The ghosts of the cat* is fun - he uncovers various regular patterns in the iterates of the cat map. Barash and Shchur [10] give an algorithm for determining certain periodic orbits in the context of using cat maps as pseudorandom number generators.

To summarize: an explicit Adler-Weiss generating partition, such as figure 14.18, completely solves the Hamiltonian cat map problem, in the sense that it generates all admissible orbits. Rational and irrational initial states generate periodic and ergodic orbits, respectively [29, 45], where every state space orbit is uniquely labeled by an admissible bi-infinite itinerary of symbols from alphabet $\mathcal{A}$ (continued in remark 15.4).

References

- [1] G. ’t Hooft and M. Veltman, “**DIAGRAMMAR**”, in *Particle Interactions at Very High Energies* (Springer, Berlin, 1974), pp. 177–322.
- [2] R. H. Abraham and C. D. Shaw, *Dynamics - The Geometry of Behavior* (Wesley, Reading, MA, 1992).
- [3] R. L. Adler and B. Weiss, “**Entropy, a complete metric invariant for automorphisms of the torus**”, *Proc. Natl. Acad. Sci. USA* **57**, 1573–1576 (1967).
- [4] R. L. Adler and B. Weiss, *Similarity of automorphisms of the torus*, Vol. 98, *Memoirs Amer. Math. Soc.* (Amer. Math. Soc., Providence RI, 1970).
- [5] D. V. Anosov, A. V. Klimenko, and G. Kolutsky, *On the hyperbolic automorphisms of the 2-torus and their Markov partitions*, 2008.

- [6] V. I. Arnol'd and A. Avez, *Ergodic Problems of Classical Mechanics* (Addison-Wesley, Redwood City, 1989).
- [7] V. Baladi, “Infinite kneading matrices and weighted zeta functions of interval maps”, *J. Functional Analysis* **128**, 226–244 (1995).
- [8] V. Baladi and D. Ruelle, “An extension of the theorem of Milnor and Thurston on the zeta functions of interval maps”, *Ergod. Theor. Dynam. Syst.* **14**, 621–632 (1994).
- [9] V. Baladi and D. Ruelle, “Sharp determinants”, *Inv. Math.* **123**, 553–574 (1996).
- [10] L. Barash and L. N. Shchur, “Periodic orbits of the ensemble of Sinai-Arnold cat maps and pseudorandom number generation”, *Phys. Rev. E* **73**, 036701 (2006).
- [11] E. Behrends, “The ghosts of the cat”, *Ergod. Theor. Dynam. Syst.* **18**, 321–330 (1998).
- [12] E. Behrends and B. Fielder, “Periods of discretized linear Anosov maps”, *Ergod. Theor. Dynam. Syst.* **18**, 331–341 (1998).
- [13] N. Bird and F. Vivaldi, “Periodic orbits of the sawtooth maps”, *Physica D* **30**, 164–176 (1988).
- [14] M. . Blank, *Discreteness and Continuity in Problems of Chaotic Dynamics* (Amer. Math. Soc., Providence RI, 1997).
- [15] M. Blank and G. Keller, “Random perturbations of chaotic dynamical systems: stability of the spectrum”, *Nonlinearity* **11**, 1351–1364 (1998).
- [16] R. Bowen, “Markov partitions for Axiom A diffeomorphisms”, *Amer. J. Math.* **92**, 725–747 (1970).
- [17] P. Collet and J.-P. Eckman, *Iterated Maps on the Interval as Dynamical Systems* (Birkhäuser, Boston, 2009).
- [18] S. C. Creagh, “Quantum zeta function for perturbed cat maps”, *Chaos* **5**, 477–493 (1995).
- [19] R. L. Devaney, *An Introduction to Chaotic Dynamical systems*, 2nd ed. (Westview Press, 2008).
- [20] F. Diacu and P. Holmes, *Celestial Encounters: The Origins of Chaos and Stability* (Princeton Univ. Press, Princeton, NJ, 1996).
- [21] D. J. Driebe, *Fully Chaotic Maps and Broken Time Symmetry* (Springer, New York, 1999).
- [22] F. J. Dyson and H. Falk, “Period of a discrete cat mapping”, *Amer. Math. Monthly* **99**, 603–614 (1992).
- [23] D. Fournier, H. Kawakami, and C. Mira, “Bifurcations of a quadratic two-dimensional diffeomorphism, homoclinic and heteroclinic situations”, *C. R. Acad. Sci., Paris* **298**, 253–256 (1984).
- [24] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer, New York, 1983).
- [25] B. Gutkin, L. Han, R. Jafari, A. K. Saremi, and P. Cvitanović, “Linear encoding of the spatiotemporal cat”, *Nonlinearity* **34**, 2800–2836 (2021).

- [26] K. T. Hansen, [Symbolic Dynamics in Chaotic systems](#), PhD thesis (Univ. of Oslo, Oslo, Norway, 1993).
- [27] S. Isola, “ [\$\zeta\$ -functions and distribution of periodic orbits of toral automorphisms](#)”, [Europhys. Lett.](#) **11**, 517–522 (1990).
- [28] J. P. Keating, “[Asymptotic properties of the periodic orbits of the cat maps](#)”, [Nonlinearity](#) **4**, 277 (1991).
- [29] J. P. Keating, “[The cat maps: quantum mechanics and classical motion](#)”, [Nonlinearity](#) **4**, 309–341 (1991).
- [30] I. G. Kevrekidis, B. Nicolaenko, and J. C. Scovel, “[Back in the saddle again: a computer assisted study of the Kuramoto-Sivashinsky equation](#)”, [SIAM J. Appl. Math.](#) **50**, 760–790 (1990).
- [31] V. Kirk, C. Postlethwaite, and A. Rucklidge, “[Resonance bifurcations of robust heteroclinic networks](#)”, [SIAM J. Appl. Dyn. Sys.](#) **11**, 1360–1401 (2012).
- [32] B. P. Kitchens, [Symbolic Dynamics: One-sided, Two-sided and Countable State Markov Shifts](#) (Springer, Berlin, 1997).
- [33] G. Knight and R. Klages, “[Linear and fractal diffusion coefficients in a family of one-dimensional chaotic maps](#)”, [Nonlinearity](#) **24**, 227 (2011).
- [34] Y. I. Manin, [Mathematics as Metaphor: Selected Essays](#) (Amer. Math. Soc., 2007).
- [35] A. Manning, “[Axiom A diffeomorphisms have rational zeta function](#)”, [Bull. London Math. Soc.](#) **3**, 215–220 (1971).
- [36] N. Metropolis, M. L. Stein, and P. R. Stein, “[On finite limit sets for transformations on the unit interval](#)”, [J. Combin. Theory](#) **15**, 25–44 (1973).
- [37] J. Milnor and W. Thurston, “[On iterated maps of the interval](#)”, in [Dynamical Systems](#), edited by J. C. Alexander (Springer, New York, 1988), pp. 465–563.
- [38] C. Mira, [Chaotic Dynamics - From One Dimensional Endomorphism to Two Dimensional Diffeomorphism](#) (World Scientific, Singapore, 1987).
- [39] P. J. Myrberg, “[Iteration der reellen Polynome zweiten Grades I](#)”, [Ann. Acad. Sc. Fenn. A](#) **256**, 1–10 (1958).
- [40] P. J. Myrberg, “[Iteration von quadratwurzeloperationen](#)”, [Ann. Acad. Sc. Fenn. A](#) **259**, 1–10 (1958).
- [41] P. J. Myrberg, “[Iteration der reellen Polynome zweiten Grades II](#)”, [Ann. Acad. Sc. Fenn. A](#) **268**, 1–13 (1959).
- [42] P. J. Myrberg, “[Sur l’itération des polynômes réels quadratiques](#)”, [J. Math. Pures Appl.](#) **41**, 339–351 (1962).
- [43] P. J. Myrberg, “[Iteration der reellen Polynome zweiten Grades III](#)”, [Ann. Acad. Sc. Fenn. A](#) **336**, 1–13 (1963).
- [44] I. Percival and F. Vivaldi, “[A linear code for the sawtooth and cat maps](#)”, [Physica D](#) **27**, 373–386 (1987).
- [45] I. Percival and F. Vivaldi, “[Arithmetical properties of strongly chaotic motions](#)”, [Physica D](#) **25**, 105–130 (1987).

- [46] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*, 3rd ed. (Cambridge Univ. Press, Cambridge UK, 2007).
- [47] C. Robinson, *Dynamical Systems: Stability, Symbolic Dynamics, and Chaos* (CRC Press, Boca Raton FL, 1995).
- [48] R. C. Robinson, *An Introduction to Dynamical Systems: Continuous and Discrete* (Amer. Math. Soc., New York, 2012).
- [49] A. Rodriguez-Gonzalez, *Qualitative Dynamics and Chaos* (2023).
- [50] A. N. Šarkovskii, “Coexistence of cycles of a continuous map of a line into itself”, *Ukrain. Math. J.* **16**, 61–71 (1964).
- [51] A. Siemaszko and M. P. Wojtkowski, “Counting Berg partitions”, *Nonlinearity* **24**, 2383–2403 (2011).

14.7 Examples

Example 14.1. 3-disk state space partition. (Continued from example 14.2) Embedded within $\mathcal{M}_{12}$, $\mathcal{M}_{13}$ are four strips $\mathcal{M}_{121}$, $\mathcal{M}_{123}$, $\mathcal{M}_{131}$, $\mathcal{M}_{132}$ of initial conditions that survive two bounces, and so forth. At each bounce a cone of initially nearby trajectories disperses (see figures 1.8 and 14.4). In order to attain a desired longer and longer itinerary of bounces, the strip of initial points $x_0 = (s_0, p_0)$ requires exponentially finer precision, nested within the initial state space strips drawn in figure 14.5. Provided that the disks are sufficiently separated, after n bounces the survivors live on 2^n exponentially thin strips, each labeled by a distinct itinerary $s_1 s_2 s_3 \dots s_n$. (continued in example 15.3)

[return to p. 262](#)

Example 14.2. 3-disk symbolic dynamics. Consider the motion of a free point particle in a plane with 3 elastically reflecting convex disks, figure 14.4. After a collision with a disk a particle either continues to another disk or escapes, so a trajectory can be labeled by the disk sequence. Sets of configuration space pinball trajectories of figure 14.4 become quickly hard to disentangle. As we shall see in what follows, their state space visualization in terms of Poincaré sections $\mathcal{P} = [s, p]$ (figure 14.5, see also figure 15.15 (b)) is much more powerful. (continued in example 14.1)

[exercise 1.1](#)

[return to p. 262](#)

Example 14.3. Rössler attractor return map: Stretch & fold. (Continued from example 4.5) In the Rössler flow (2.30) of example 3.2 we sketched the attractor by running a long chaotic trajectory, and noted that the attractor of figure 14.7 (a) is very thin. For Rössler flow an interval transverse to the attractor is stretched, folded and fiercely pressed back. The attractor is ‘fractal’, but for all practical purposes the return map is 1-dimensional; your printer will need a resolution better than 10^{13} dots per inch to start resolving its structure. We had attempted to describe this ‘stretch & fold’ flow by a 1-dimensional return map, but the maps that we plotted in figure 3.4 were disquieting; they did not appear to be a 1-to-1 maps. This apparent non-invertibility is an artifact of projection of a 2-dimensional return map $(R_n, z_n) \rightarrow (R_{n+1}, z_{n+1})$ onto the 1-dimensional subspace $R_n \rightarrow R_{n+1}$. Now that we understand equilibria and their linear stability, let’s do this right.

The key idea is to measure *arclength* distances along the unstable manifold of the x_- equilibrium point, as in figure 14.7 (a) (arclength parametrization of unstable manifolds is discussed in detail in sect. 15.1.1). Luck is with us; figure 14.7 (b) return map $s_{n+1} = P(s_n)$ looks much like a parabola of example 3.7, so we shall take the unimodal map symbolic dynamics, sect. 14.3, as our guess for the covering symbolic dynamics.

(continued in example 14.11)

[return to p. 265](#)

Example 14.4. Lorenz flow: Stretch & crease. We now deploy the symmetry of Lorenz flow to streamline and complete analysis of the Lorenz strange attractor commenced in example 11.8. There we showed that the rotational $C_2 = \{e, d\}$ symmetry identifies the two equilibria EQ_1 and EQ_2 , and the traditional ‘two-eared’ Lorenz flow figure 2.5 is replaced by the ‘single-eared’ flow of figure 11.2 (b). Furthermore, the C_2 symmetry identifies the two half-planes of any plane through the z axis, replacing a full-space Poincaré section plane by a half-plane, and the two directions of a full-space eigenvector of EQ_0 by a one-sided eigenvector, see figure 11.2 (b).

Example 4.7 explained the genesis of the x_{EQ_1} equilibrium unstable manifold, its orientation and thickness, its collision with the z -axis, and its heteroclinic connection to

Figure 14.14: (a) A Poincaré section of the Lorenz flow in the doubled-polar angle representation, figure 11.2 (b), given by the $[y', z]$ plane that contains the z -axis and the equilibrium EQ_1 . Most of the section plane except for the two shaded trapezoids is removed to aid visualization of the flow. x' axis points toward the viewer. (b) The Poincaré section plane. Crossings *into* the section are marked red (solid) and crossings *out of* the section are marked blue (dashed). Outermost points of both in- and out-sections are given by the EQ_0 unstable manifold $W^u(EQ_0)$ intersections. (E. Siminos)

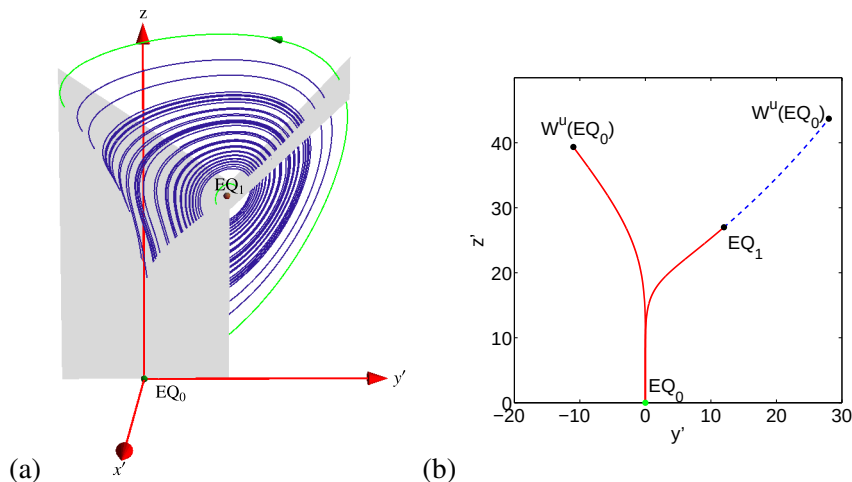
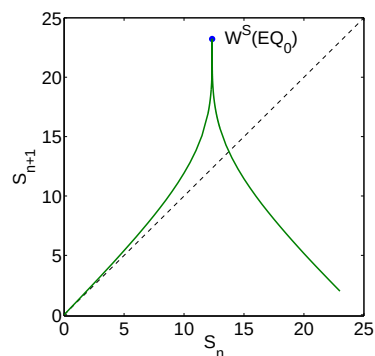


Figure 14.15: The return map $s_{n+1} = P(s_n)$ parameterized by Euclidean arclength s measured along the EQ_1 unstable manifold, from x_{EQ_1} to $W^u(EQ_0)$ section point, uppermost right point of the blue (dashed) segment in figure 14.14 (b). The critical point (the ‘crease’) of the map is given by the section of the heteroclinic orbit $W^s(EQ_0)$ that descends all the way to EQ_0 , in infinite time and with infinite slope. (E. Siminos)



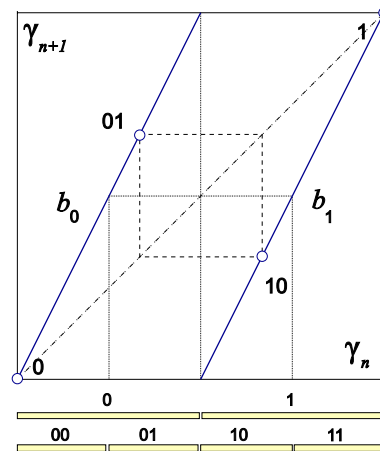
the $x_{EQ_0} = (0, 0, 0)$ equilibrium. All that remains is to describe how the EQ_0 neighborhood connects back to the EQ_1 unstable manifold.

Figure 11.2 (b) and figure 14.14 (a) show clearly how the Lorenz dynamics is pieced together from the 2 equilibria and their unstable manifolds: Having completed the descent to EQ_0 , the infinitesimal neighborhood of the heteroclinic $EQ_1 \rightarrow EQ_0$ trajectory is ejected along the unstable manifold of EQ_0 and is re-injected into the unstable manifold of EQ_1 . Both sides of the narrow strip enclosing the EQ_0 unstable manifold lie above it, and they get folded onto each other with a knife-edge crease (contracted exponentially for infinite time to the EQ_0 heteroclinic point), with the heteroclinic out-trajectory defining the outer edge of the strange attractor. This leads to the folding of the outer branch of the Lorenz strange attractor, illustrated in figure 14.14 (b), with the outermost edge following the unstable manifold of EQ_0 .

Now the stage is set for construction of Poincaré sections and associated return maps. There are two natural choices; the section at EQ_0 , lower part of figure 14.14 (b), and the section (blue) above EQ_1 . The first section, together with the blowup of the EQ_0 neighborhood, figure 4.6 (b), illustrates clearly the scarcity of trajectories (vanishing natural measure) in the neighborhood of EQ_0 . The flat section above EQ_1 (which is, believe it or not, a smooth conjugacy by the flow of the knife-sharp section at EQ_0) is more convenient for our purposes. Its return map (3.4) is given by figure 14.15.

The rest is straight sailing: to accuracy 10^{-4} the return map is unimodal, its critical point's forward trajectory yields the kneading sequence (14.5), and the admissible binary sequences, so any number of periodic points can be accurately determined from this 1-dimensional return map, and the 3-dimensional cycles then verified by integrating the

Figure 14.16: The $n = 2$ and 4 intervals state space partitions for the Bernoulli shift map (14.18), together with the fixed points $\bar{0}$, $\bar{1}$ and the 2-cycle $\bar{01}$. For the doubling map (14.19) acting on the circle, $\bar{0}$ is the only one fixed point, as $x_1 = 1 = 0 \pmod{1}$.



Lorenz differential equations (2.25). As already observed by Lorenz, such a map is everywhere expanding on the strange attractor, so it is no wonder mathematicians can here make the ergodicity rigorous.

(E. Siminos and J. Halcrow)

section 23.7
return to p. 265

Example 14.5. Bernoulli shift map state space partition. First, an easy example: the *Bernoulli shift map*, figure 14.16,

$$b(\gamma) = \begin{cases} b_0(\gamma) = 2\gamma, & \gamma \in \mathcal{M}_0 = [0, 1/2) \\ b_1(\gamma) = 2\gamma - 1, & \gamma \in \mathcal{M}_1 = (1/2, 1] \end{cases}, \quad (14.18)$$

models the 50-50% probability of a coin toss. It maps the unit interval onto itself, with fixed points $\gamma_0 = 0$, $\gamma_1 = 1$. The closely related *doubling map* acts on the circle

$$\gamma \mapsto 2\gamma \pmod{1}, \quad \gamma \in [0, 1) \quad (14.19)$$

and consequently has only one fixed point, $\gamma_0 = 0 = 1 \pmod{1}$. The Bernoulli map is called a ‘shift’ map, as a multiplication by 2 acts on the binary representation of $\gamma = .s_1s_2s_3\cdots$ by shifting its digits, $b(\gamma) = .s_2s_3\cdots$. The n th preimages $b^{-(n-1)}(\gamma)$ of the critical point $\gamma_c = 1/2$ partition the state space into 2^n subintervals, each labeled by the first n binary digits of points $\gamma = .s_1s_2s_3\cdots$ within the subinterval: figure 14.16 illustrates such 4-intervals state space partition $\{\mathcal{M}_{00}, \mathcal{M}_{01}, \mathcal{M}_{11}, \mathcal{M}_{10}\}$ for $n = 2$.

Consider a map $f(x)$ topologically conjugate (two monotonically increasing branches) to the Bernoulli shift, with the forward orbit of x generating the itinerary $s_1s_2s_3\cdots$. Convert this itinerary into Bernoulli map point $\gamma = .s_1s_2s_3\cdots$. These values can now be used to *spatially order* points with different temporal itineraries: if $\gamma < \gamma'$, then $x < x'$.

Suppose we have already computed all $(n-1)$ -cycles of $f(x)$, and would now like to compute the cycle $p = s_1s_2s_3\cdots s_n$ of period n . Mark γ values on the unit interval for all known periodic points of the Bernoulli shift map, and then insert in between them $\gamma_{\sigma^k p}, k = 0, 1, \cdots, n_p - 1$ corresponding to periodic points of cycle p . In the dynamical state space they will be bracketed by corresponding cycle points x_j from cycles already computed, and thus the knowledge of the topological ordering of *all* cycle points provides us with robust initial guesses for periodic-orbit searches for any map with 2 monotonically increasing branches. (continued in example 28.2)

return to p. 266

Example 14.6. Unimodal maps. (Continued from example 3.7) The simplest examples of unimodal maps are the quadratic map (3.21)

$$f(x) = Ax(1-x), \quad x \in \mathcal{M} = [0, 1] \quad (14.20)$$

and numerically computed return maps such as figure 14.7 (b). The two forms (3.21) and (14.20) of the quadratic map are related by

$$x \rightarrow \frac{x - \frac{1}{2}}{\frac{A}{4} - \frac{1}{2}}, \quad a = A \left(\frac{A}{4} - \frac{1}{2} \right).$$

Such dynamical systems are irreversible (the inverse of f is double-valued), but, as we shall show in sect. 15.2, they may nevertheless serve as effective descriptions of invertible 2-dimensional hyperbolic flows. For the unimodal map such as figure 14.9 a Markov partition of the unit interval $\mathcal{M}$ is given by the two intervals $\{\mathcal{M}_0, \mathcal{M}_1\}$. (continued in example 14.7)

[return to p. 267](#)

Example 14.7. Full tent map, Ulam map. (Continued from example 14.6) The simplest examples of unimodal maps with complete binary symbolic dynamics are the *full tent map*, figure 14.9 (a),

$$f(\gamma) = 1 - 2|\gamma - 1/2|, \quad \gamma \in \mathcal{M} = [0, 1], \quad (14.21)$$

the Ulam map (quadratic map (14.20) with $A = 4$)

[exercise A2.3](#)

$$f(x) = 4x(1 - x), \quad x \in \mathcal{M} = [0, 1], \quad (14.22)$$

and the repelling unimodal maps such as figure 14.9. For unimodal maps the Markov partition of the unit interval $\mathcal{M}$ is given by intervals $\{\mathcal{M}_0, \mathcal{M}_1\}$. We refer to (14.21) as the *complete* tent map because its symbolic dynamics is completely binary: as both $f(\mathcal{M}_0)$ and $f(\mathcal{M}_1)$ fully cover $\mathcal{M} = \{\mathcal{M}_0, \mathcal{M}_1\}$, all binary sequences are realized as admissible itineraries.

[return to p. 267](#)

Example 14.8. Periodic orbits of unimodal maps. Let

$$f(x) = \begin{cases} f_0(x) & \text{if } x < x_c \\ f_1(x) & \text{if } x > x_c \end{cases}, \quad (14.23)$$

and assume that all periodic orbits are unstable, i.e., the stability $\Lambda_p = f_a^{k'}$ (see (4.43)) satisfies $|\Lambda_p| > 1$. Then the periodic point $x_{s_0 s_1 s_2 \dots s_{n-1}}$ is the only fixed point of the unique composition (3.14) of n maps

$$f_{s_n} \circ \dots \circ f_{s_2} \circ f_{s_1}(x_{s_0 s_1 s_2 \dots s_{n-1}}) = x_{s_0 s_1 s_2 \dots s_{n-1}} \quad (14.24)$$

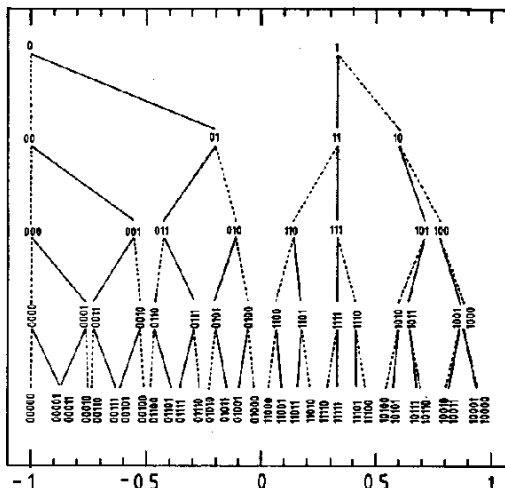
(note that successive maps, applied from the left, correspond to later times, i.e., later symbols in the itinerary).

The n th iterate of a unimodal map has at most 2^n monotone segments, and therefore there will be 2^n or fewer periodic points of period n . For the full tent map (14.21) it has exactly 2^n periodic points. A periodic orbit p of period n corresponds to an infinite repetition of a period $n = n_p$ symbol string block, customarily indicated by a line over the string: $p = S_p = (s_1 s_2 s_3 \dots s_n)^\infty = \overline{s_1 s_2 s_3 \dots s_n}$. As all itineraries are infinite, we shall adopt convention that a finite string itinerary $p = s_1 s_2 s_3 \dots s_n$ stands for infinite repetition of a finite block, and routinely omit the overline. A cycle p is called *prime* if its itinerary S cannot be written as a repetition of a shorter block S' . If the itinerary of x_0 is $p = s_1 s_2 s_3 \dots s_n$, its cyclic permutation $\sigma^k p = \overline{s_k s_{k+1} \dots s_n s_1 \dots s_{k-1}}$ corresponds to the point x_{k-1} in the same cycle.

[section 15.2](#)

[return to p. 269](#)

Figure 14.17: The alternating binary tree organization of the periodic points of a unimodal map (in this example, the Ulam map (14.22)). The itinerary of a point is read off the tree by starting at the root and following the branches down to x ; relative ordering of points along the x axis is given by the relative ordering of the corresponding nodes.



Example 14.9. Systematic searches for unimodal map cycles. Knowledge of the topological coordinate (14.4) is very useful when searching for periodic orbits. Assume that we have already determined all periodic points $x_a, x_b, \dots$ of period n , and would like to have a good initial guess for the period $(n+1)$ periodic point x_d with prescribed itinerary $S_d^+ := S^+(x_d)$. It is easy to determine the two closest itineraries $\gamma(S_a^+)$, $\gamma(S_b^+)$ that bracket $\gamma(S_d^+)$. If $\gamma(S_a^+) < \gamma(S_d^+) < \gamma(S_b^+)$, then one can restrict the search for x_c to $x_c \in [x_a, x_b]$. For example, the relative ordering of all unimodal map periodic points up to $n = 5$ is given in the figure 14.17. Appendix A14.1 contains further details of the symbolics dynamics for periodic point of unimodal maps.

[return to p. 269](#)

Example 14.10. Periodic points of the full tent map. Each cycle p is a set of n_p rational-valued full tent map periodic points γ . It follows from (14.4) that if the repeating string $s_1 s_2 \dots s_n$ contains an odd number of '1's, the string of well ordered symbols $w_1 w_2 \dots w_{2n}$ has to be of the double length before it repeats itself. The cycle-point γ is a geometrical sum which we can rewrite as the odd-denominator fraction

$$\begin{aligned} \gamma(\overline{s_1 s_2 \dots s_n}) &= \sum_{t=1}^{2n} \frac{w_t}{2^t} + \frac{1}{2^{2n}} \sum_{t=1}^{2n} \frac{w_t}{2^t} + \dots \\ &= \frac{2^{2n}}{2^{2n} - 1} \sum_{t=1}^{2n} \frac{w_t}{2^t} \end{aligned} \quad (14.25)$$

Using this we can calculate the $\hat{\gamma}_p = \hat{\gamma}(S_p)$ for all short cycles. For orbits up to period 5 this is done in table 14.1.

[return to p. 269](#)

Example 14.11. Rössler return map web diagram. (Continuation of example 14.1) The arclength distance along the unstable manifold of the x_- equilibrium point return map, figure 14.7 (b), generates the kneading sequence (14.5) as the itinerary of the critical point plotted in figure 14.13 (a).

[return to p. 272](#)

Example 14.12. Adler-Weiss partition of the cat map state space. (Continued from example 8.9) Cat maps, also known as Thom-Anosov diffeomorphisms, or Thom-Anosov-Arnol'd-Sinai cat maps, have been extensively studied as the simplest examples of chaotic Hamiltonian systems.

[remark 14.5](#)

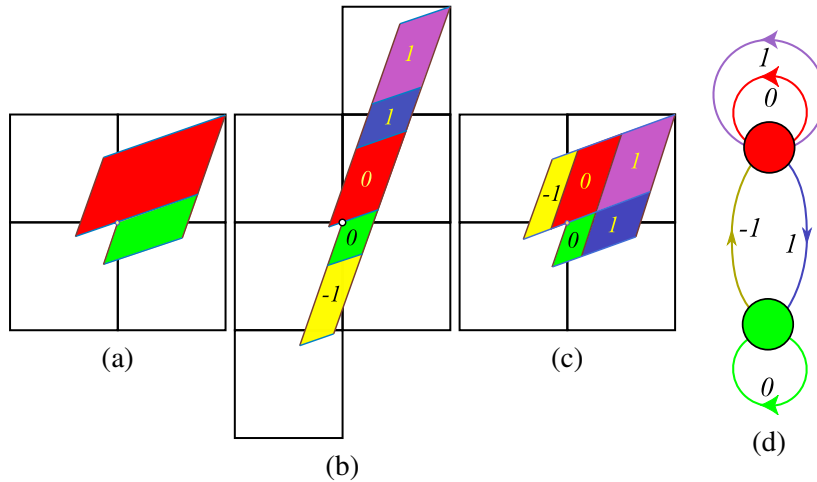


Figure 14.18: (a) An Adler-Weiss generating partition of the unit torus for the $s = 3$ Percival-Vivaldi cat map (8.52), with rectangle M_A (red) and M_B (green) borders given by the cat map stable (blue) and unstable (dark red) manifolds, i.e., along the two eigenvectors corresponding to the eigenvalues (8.10). (b) Mapped one step forward in time, the rectangles are stretched along the unstable direction and shrunk along the stable direction. Sub-rectangles M_j that have to be translated back into the partition are indicated by color and labeled by their lattice translation $m_j \in \mathcal{A} = \{\underline{1}, 0, 1\}$, which also doubles as the 3-letter alphabet $\mathcal{A}$. (c) The sub-rectangles M_j translated back into the initial partition yield a generating partition, with the finite grammar given by the transition graph (d). The nodes refer to the rectangles A and B , and the five links correspond to the five sub-rectangles induced by one step forward-time dynamics (continued in figure 18.4)

Percival-Vivaldi cat map (8.52) is a discrete time non-autonomous Hamiltonian system, time-forced by ‘pulses’ m_t . The m_t translations reshuffle the state space, as in figure 14.18, thus partitioning it into regions M_m , labeled with letters m of the $|\mathcal{A}|$ -letter alphabet $\mathcal{A}$, and associating a symbol sequence $\{m_t\}$ to the dynamical trajectory $\{x_t\}$. As the relation (8.52) between the trajectory x_t and its symbolic dynamics encoding m_t is linear, Percival and Vivaldi refer to m_t as a ‘linear code’. The deep problem with the Percival-Vivaldi code prescription is that it does not yield a generating partition; the borders (i.e., x_0, x_1 axes) of their unit-square partition $(x_{t-1}, x_t) \in (0, 1] \times (0, 1]$ do not map onto themselves, resulting in the infinity of, to us unknown, grammar rules for inadmissible symbol sequences.

This problem is resolved by the Adler-Weiss partition which utilizes the stable/unstable manifolds of the fixed point at the origin to cover a unit area torus by a two-rectangles generating partition; for the Percival-Vivaldi cat map (8.52), such partition is drawn in figure 14.18. Symbolic dynamics on this partition is a subshift of finite type, with the 3-letter alphabet

$$\mathcal{A} = \{\underline{1}, 0, 1\} \quad (14.26)$$

that indicates the translation needed to return the given sub-rectangle M_j back into the two-rectangle partition $M = M_A \cup M_B$. We have introduced here the symbol $\underline{m}_n$ to denote m_n with the negative sign, i.e., ‘ $\underline{1}$ ’ stands for symbol ‘ -1 ’. The effect of choosing the dynamically intrinsic, two-rectangles cover of figure 14.18 is magic - now every infinite walk on the transition graph of figure 14.18 (d) corresponds to a unique admissible orbit $\{x_t\}$, and the transition graph generates all admissible itineraries $\{m_t\}$ (continued in example 18.13).

(H. Liang and P. Cvitanović)

return to p. 275

Exercises

- 14.1. **Binary symbolic dynamics.** Verify that the shortest prime binary cycles of the unimodal repeller of figure 14.9 are $\overline{0}$, $\overline{1}$, $\overline{01}$, $\overline{001}$, $\overline{011}$, $\dots$. Compare with table 18.1. Sketch them in the graph of the unimodal function $f(x)$; compare the ordering of the periodic points with that in figure 14.10. The point is that while overlayed on each other the longer cycles look like a hopeless jumble, the periodic points are clearly and logically ordered by the alternating binary tree.
- 14.2. **Generating prime cycles.** Write a program that generates all binary prime cycles up to a given finite period.
- 14.3. **A contracting baker's map.** Consider the contracting (or “dissipative”) baker's map defined in exercise 4.6.
The symbolic dynamics encoding of trajectories is realized via symbols 0 ($y \leq 1/2$) and 1 ($y > 1/2$). Consider the observable $a(x, y) = x$. Verify that for any periodic orbit $p = \overline{s_1 \dots s_{n_p}}$, $s_i \in \{0, 1\}$ the Birkhoff sum is

$$A_p = \frac{3}{4} \sum_{j=1}^{n_p} \delta_{s_j, 1}.$$

For the definition of a Birkhoff sum, see sect. 20.1.1.

- 14.4. **Unimodal map symbolic dynamics.** Show that the tent map point $\gamma(S^+)$ with future itinerary S^+ is given by converting the sequence of s_n 's into a binary number by the algorithm (14.4). This follows by inspection from the binary tree of figure 14.10.
- 14.5. **Unimodal map kneading value.** Consider the 1-dimensional quadratic map

$$f(x) = Ax(1 - x), \quad A = 3.8. \quad (14.27)$$

- (a) (easy) Plot (14.27), and the first 4-8 (whatever looks better) iterates of the critical point $x_c = 1/2$.
- (b) (hard) Draw corresponding intervals of the partition of the unit interval as levels of a Cantor set, as in the symbolic dynamics partition of figure 14.9. Note, however, that some of the intervals of figure 14.9 do not appear in this case - they are *pruned*.
- (c) (easy) Check numerically that $K = S^+(x_c)$, kneading sequence (the itinerary of the critical point (14.5)) is

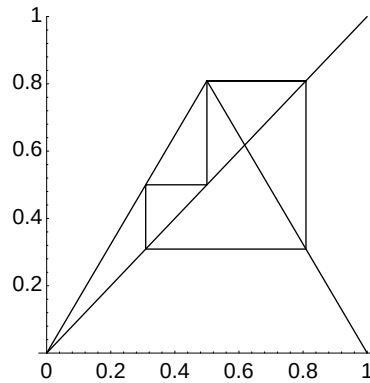
$$K = 101101111011011110101111011110\dots$$

As the orbits of a chaotic map are exponentially unstable, so many digits seem too good to be true - recheck this sequence using arbitrary precision arithmetics.

- (d) (medium) The tent map point $\gamma(S^+)$ with future itinerary S^+ is given by converting the sequence of s_n 's into a binary number by the algorithm (14.4). List the corresponding kneading value (14.5) sequence $\kappa = \gamma(K)$ to the same number of digits as K .
- (e) (hard) Plot the dike map, figure 14.11, with the same kneading sequence K as $f(x)$. The dike map is obtained by slicing off all $\gamma(S^+(x_0)) > \kappa$, from the full tent map figure 14.9 (a), see (14.6).

How this kneading sequence is converted into a series of pruning rules is a dark art, relegated to sect. 18.5.

- 14.6. **‘Golden mean’ pruned map.** Consider a symmetric tent map on the unit interval such that its highest point belongs to a 3-cycle:



- Find the value $|\Lambda|$ for the slope (the two different slopes $\pm\Lambda$ just differ by a sign) where the maximum at $1/2$ is a periodic point in a 3-cycle, as depicted in the figure.
- Show that no orbit of this map can visit the region $x > (1 + \sqrt{5})/4$ more than once. Verify also that once an orbit exceeds $x > (\sqrt{5} - 1)/4$, it does not reenter the region $x < (\sqrt{5} - 1)/4$.
- If an orbit is in the interval $(\sqrt{5} - 1)/4 < x < 1/2$, where will it be on the next iteration?
- If the symbolic dynamics is such that for $x < 1/2$ we use the symbol 0 and for $x > 1/2$ we use the symbol 1, show that no periodic orbit will have the substring `_00_` in it.
- On a second thought, is there a periodic orbit that violates the above `_00_` pruning rule?

The ‘golden mean’ map is the simplest example of a system whose symbolic dynamics is defined by a finite set of pruning rules (in this case one pruned block). We shall revisit it in exercise 18.7, exercise 22.1, exercise 18.6 and exercise 18.8.

- 14.7. **Binary 3-step transition matrix.** Construct an $[8 \times 8]$ binary 3-step transition matrix analogous to the 2-step transition matrix (17.11). Convince yourself that the number of terms of contributing to $\text{tr } T^n$ is independent of the memory length, and that this $[2^m \times 2^m]$ trace is well defined in the infinite memory limit $m \rightarrow \infty$.
- 14.8. **Full tent map periodic points.** This exercise is easy: just making sure you know how to go back and forth between spatial and temporal ordering of trajectory points.
- compute the two periodic points of cycle $\overline{01}$ “by hand,” by solving the fixed-point condition for the second iterate $f_1 \circ f_0$
 - compute the periodic points of two 3-cycles $\overline{001}$ and $\overline{011}$ by solving the fixed-point condition for the third iterates
 - compute the five periodic points of cycle $\overline{10011}$ using (14.25)
 - compute the five periodic points of cycle $\overline{10000}$
 - derive (14.25)
 - (optional) plot the above two 5-cycles on the graph of the full tent map, and as many others as you find interesting. Why? Because you can start appreciating the power of kneading theory—while the state space orbits get more and more complicated and impenetrable, the kneading sequence pruning rule is as simple and as sharp as a knife.

(continued in exercise 16.1)

Chapter 15

Stretch, fold, prune

I.1. Introduction to conjugacy problems for diffeomorphisms. This is a survey article on the area of global analysis defined by differentiable dynamical systems or equivalently the action (differentiable) of a Lie group G on a manifold M . Here $\text{Diff}(M)$ is the group of all diffeomorphisms of M and a diffeomorphism is a differentiable map with a differentiable inverse. (...) Our problem is to study the global structure, i.e., all of the orbits of M .

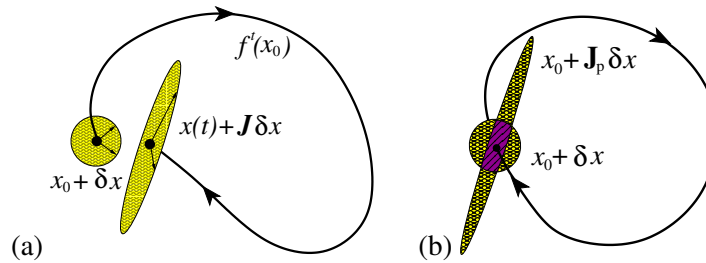
—Stephen Smale, *Differentiable Dynamical Systems*

WE HAVE LEARNED that the Rössler attractor is very thin, but otherwise the return maps that we found were disquieting – figure 3.4 did not appear to be a one-to-one map. This apparent loss of invertibility is an artifact of projection of higher-dimensional return maps onto their lower-dimensional subspaces. As the choice of a lower-dimensional subspace is arbitrary, the resulting snapshots of return maps look rather arbitrary, too. Such observations beg a question: Does there exist a natural, intrinsic coordinate system in which we should plot a return map?

We shall argue in sect. 15.1 that the answer is yes. The intrinsic coordinates are given by the stable/unstable manifolds, and a return map should be plotted as a map from the unstable manifold back onto the immediate neighborhood of the unstable manifold. In chapter 5, we established that Floquet multipliers of periodic orbits are (local) dynamical invariants. Here we shall show that every equilibrium point and every periodic orbit carries with it stable and unstable manifolds which provide topologically invariant *global* foliation of the state space. They will enable us to partition the state space in a dynamically invariant way and assign symbolic dynamics itineraries to trajectories.

The topology of stretching and folding fixes the relative spatial ordering of trajectories, and separates the admissible and inadmissible itineraries. We illustrate how this works on Hénon map example 15.3. Determining which symbol sequences are absent, or ‘pruned’ is a formidable problem when viewed in the state

Figure 15.1: (a) An infinitesimal neighborhood transported along a trajectory $x(t)$. (b) An infinitesimal Poincaré section neighborhood brought back in one period.



space, $[x_1, x_2, \dots, x_d]$ coordinates. It is equivalent to the problem of determining the location of all homoclinic tangencies, or all turning points of the Hénon attractor. They are dense on the attractor and show no self-similar structure in the state space coordinates. However, in the ‘danish pastry’ representation of sect. 15.3 (and the ‘pruned danish’, in American vernacular, of sect. 15.4), the pruning problem is visualized as crisply as the New York subway map; any itinerary which strays into the ‘pruned region’ is banned.

The level is distinctly cyclist, in distinction to the pedestrian tempo of the preceding chapter. Skip most of this chapter unless you really need to get into nitty-gritty details of symbolic dynamics.



fast track:
chapter 16, p. 321

15.1 Goin’ global: stable/unstable manifolds

The complexity of this figure will be striking, and I shall not even try to draw it.

— H. Poincaré, on his discovery of homoclinic tangles, *Les méthodes nouvelles de la mécanique céleste* [37]

The Jacobian matrix J^t transports an infinitesimal neighborhood. The eigenvalues and eigen-directions of J^t transforms an initial infinitesimal frame of neighboring trajectories into a distorted frame time t later, as in figure 15.1 (a) and figure 4.1. Nearby trajectories separate exponentially along the unstable directions, approach each other along the stable directions, and creep along the marginal directions.

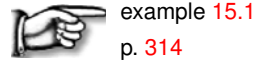
The Poincaré section fixed point q (as in figure 15.1 (b)) Jacobian matrix $J(x)$ eigenvectors (5.10) form a rectilinear coordinate frame in which the flow into, out of, or encircling the fixed point is linear in the sense of sect. 4.3.

The continuations of the span of the local stable, unstable eigen-directions into global curvilinear invariant manifolds are called the *stable*, respectively *unstable manifolds*. They consist of all points which march into the fixed point forward,

respectively backward in time

$$\begin{aligned} W^s &= \{x \in \mathcal{M} : f^t(x) - x_q \rightarrow 0 \text{ as } t \rightarrow \infty\} \\ W^u &= \{x \in \mathcal{M} : f^{-t}(x) - x_q \rightarrow 0 \text{ as } t \rightarrow \infty\}. \end{aligned} \quad (15.1)$$

Eigenvectors $\mathbf{e}^{(i)}$ of the monodromy matrix $J(x)$ play a special role - on them the action of the dynamics is the linear multiplication by Λ_i (for a real eigenvector) along 1-dimensional invariant curve $W_{(i)}^{u,s}$ or spiral in/out action in a 2- D surface (for a complex pair). For $t \rightarrow \pm\infty$ a finite segment on $W_{(c)}^s$, respectively $W_{(e)}^u$ converges to the linearized map eigenvector $\mathbf{e}^{(c)}$, respectively $\mathbf{e}^{(e)}$, where $^{(c)}$, $^{(e)}$ stand respectively for ‘contracting’, ‘expanding.’ In this sense each eigenvector defines a (curvilinear) axis of the stable, respectively unstable manifold.



example 15.1
p. 314

Actual construction of these manifolds is the converse of their definition (15.1); one starts with an arbitrarily small segment of a fixed point eigenvector and lets evolution stretch it into a finite segment of the associated manifold. As a periodic point x on cycle p is a fixed point of $f^{T_p}(x)$, the fixed point discussion that follows applies equally well to equilibria and periodic orbits.

Expanding real and positive Floquet multiplier. Consider i th expanding eigenvalue, eigenvector pair $(\Lambda_i, \mathbf{e}^{(i)})$ computed from $J = J_p(x)$ evaluated at a fixed point x ,

$$J(x)\mathbf{e}^{(i)}(x) = \Lambda_i\mathbf{e}^{(i)}(x), \quad x \in \mathcal{M}_p, \quad \Lambda_i > 1. \quad (15.2)$$

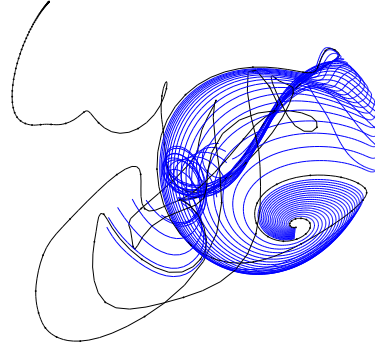
Take an infinitesimal eigenvector $\mathbf{e}^{(i)}(x)$, $\|\mathbf{e}^{(i)}(x)\| = \varepsilon \ll 1$ and its return $\Lambda_i\mathbf{e}^{(i)}(x)$ after one period T_p . Then strapprinkle the straight interval between $[\varepsilon, \Lambda_i\varepsilon] \subset W_{(i)}^u$ with a large number of points $x^{(k)}$, for example equidistantly spaced on logarithmic scale between $\ln \varepsilon$ and $\ln \Lambda_i + \ln \varepsilon$. The successive returns of these points $f^{T_p}(x^{(k)})$, $f^{2T_p}(x^{(k)})$, $\dots$, $f^{mT_p}(x^{(k)})$ trace out the 1d curve $W_{(i)}^u$ within the unstable manifold. As separations between points tend to grow exponentially, every so often one needs to interpolate new starting points between the rarified ones. Repeat for $-\mathbf{e}^{(i)}(x)$.

Contracting real and positive Floquet multiplier. Reverse the action of the map backwards in time. This turns a contracting direction into an expanding one, tracing out the curvilinear stable manifold $W_{(i)}^s$ as a continuation of $\mathbf{e}^{(i)}$.

Expanding/contracting real negative Floquet multiplier. As above, but every even iterate $f^{2T_p}(x^{(k)})$, $f^{4T_p}(x^{(k)})$, $f^{6T_p}(x^{(k)})$ continues in the direction $\mathbf{e}^{(i)}$, every odd one in the direction $-\mathbf{e}^{(i)}$.

Complex Floquet multiplier pair, expanding/contracting. The complex Floquet multiplier pair $\{\Lambda_j, \Lambda_{j+1} = \Lambda_j^*\}$ has Floquet exponents (4.8) of form $\lambda^{(j)} = \mu^{(j)} \pm i\omega^{(j)}$, with the sign of $\mu^{(k,j)} \neq 0$ determining whether the linear neighborhood is out / in spiralling. The orthogonal pair of real eigenvectors $\{\text{Re } \mathbf{e}^{(j)}, \text{Im } \mathbf{e}^{(j)}\}$

Figure 15.2: A $2d$ unstable manifold obtained by continuation from the linearized neighborhood of a complex eigenvalue pair of an unstable equilibrium of plane Couette flow, a projection from a 61,506-dimensional state space ODE truncation of the (∞ -dimensional) Navier-Stokes PDE. (J.F. Gibson, 8 Nov. 2005 blog entry [17])



spans a plane. $T = 2\pi/\omega^{(j)}$ is the time of one turn of the spiral, $J^T \text{Re } \mathbf{e}^{(j)}(x) = |\Lambda_j| \text{Re } \mathbf{e}^{(j)}(x)$. As in the real cases above, sprinkle the straight interval between $[\varepsilon, |\Lambda_j|\varepsilon]$ along $\text{Re } \mathbf{e}^{(j)}(x)$ with a large number of points $x^{(k)}$. The flow will now trace out the $2d$ invariant manifold as an out / in spiralling strip. Two low-dimensional examples are the unstable manifolds of the Lorenz flow, figure 14.14 (a), and the Rössler flow, figure 14.7 (a). For a highly non-trivial example, see figure 15.2.

The unstable manifolds of a flow are d_u -dimensional. Taken together with the marginally stable direction along the flow, they are rather hard to visualize. A more insightful visualization is offered by $(d-1)$ -dimensional Poincaré sections (3.2) with the marginal flow direction eliminated (see also sect. 3.1.2). Stable, unstable manifolds for maps are defined by

$$\begin{aligned}\hat{W}^s &= \{x \in \mathcal{P} : P^n(x) - x_q \rightarrow 0 \text{ as } n \rightarrow \infty\} \\ \hat{W}^u &= \{x \in \mathcal{P} : P^{-n}(x) - x_q \rightarrow 0 \text{ as } n \rightarrow \infty\},\end{aligned}\tag{15.3}$$

where $P(x)$ is the $(d-1)$ -dimensional return map (3.1). In what follows, all invariant manifolds W^u, W^s will be restricted to their Poincaré sections $\hat{W}^u, \hat{W}^s$.

 **example 15.2**
p. 314

In general the full state space eigenvectors do not lie in a Poincaré section; the eigenvectors $\hat{\mathbf{e}}^{(j)}$ tangent to the section are given by (5.22). Furthermore, while in the linear neighborhood of fixed point x , the trajectories return with approximate periodicity T_p . This is not the case for globally continued manifolds; $\tau(x)$ or the first return times (3.1) differ, and the $\hat{W}_{(j)}^u$ restricted to the Poincaré section is obtained by continuing trajectories of the points from the full state space curve $W_{(j)}^u$ to the section $\mathcal{P}$.

For long times the unstable manifolds wander throughout the connected ergodic component and are no more informative than an ergodic trajectory. For example, the line with equitemporal knots in figure 15.2 starts out on a smoothly curved neighborhood of the equilibrium, but only after a ‘turbulent’ episode decays into an attractive equilibrium point. The trick is to stop continuing an invariant manifold while the going is still good.



fast track:
sect. 15.2, p. 295

Learning where to stop is a bit of a technical exercise. The reader might prefer to skip the next section on a first reading.

15.1.1 Parametrization of invariant manifolds



As the flow is nonlinear, there is no ‘natural’ linear basis to represent it. Wistful hopes like ‘POD modes’, ‘Karhunen-Loève’, and other linear changes of bases do not cut it. The invariant manifolds are curved, and their coordinatizations are of necessity curvilinear, just as the maps of our globe are, but infinitely foliated and thus much harder to chart.

Let us illustrate this by parameterizing a $1d$ slice of an unstable manifold by its arclength. Sprinkle points evenly $\{x^{(1)}, x^{(2)}, \dots, x^{(N-1)}\}$ between the equilibrium point $x_q = x^{(0)}$ and point $x = x^{(N)}$, along the $1d$ unstable manifold continuation $x^{(k)} \in \hat{W}_{(j)}^u$ of the unstable $\hat{e}^{(j)}$ eigendirection (we shall omit the eigendirection label $_{(j)}$ in what follows). Then the arclength from equilibrium point $x_q = x^{(0)}$ to $x = x^{(N)}$ is given by

$$s = \lim_{N \rightarrow \infty} \sum_{k=1}^N \left(g_{ij} dx_i^{(k)} dx_j^{(k)} \right)^{1/2}, \quad dx_i^{(k)} = x_i^{(k)} - x_i^{(k-1)}. \quad (15.4)$$

For the lack of a better idea (perhaps the dynamically determined $g = J^T J$ would be a more natural metric?), let us measure arclength in the Euclidean metric, $g_{ij} = \delta_{ij}$, so

$$s = \lim_{N \rightarrow \infty} \sum_{k=1}^N \left(dx^{(k)} \cdot dx^{(k)} \right)^{1/2}. \quad (15.5)$$

By definition $f^{\tau(x)}(x) \in \hat{W}_{(j)}^u$, so $f^t(x)$ induces a $1d$ map $s(s_0, \tau) = s(f^{\tau(x_0)}(x_0))$.

Turning points are points on the unstable manifold for which the local unstable manifold curvature diverges for forward iterates of the map, i.e., points at which the manifold folds back onto itself arbitrarily sharply. For our purposes, approximate turning points suffice. The $1d$ curve $\hat{W}_{(j)}^u$ starts out linear at x_q , then gently curves until –under the influence of other unstable equilibria and/or periodic orbits– it folds back sharply at ‘turning points’ and then nearly retraces itself. This is likely to happen if there is only one unstable direction, as we saw in the Rössler attractor example 14.3, but if there are several, the ‘turning point’ might get stretched out in the non-leading expanding directions.

The trick is to figure out a good *base segment* to the nearest turning point $L = [0, s_b]$, and after the foldback assign to $s(x, t) > s_b$ the nearest point s on the base segment. If the stable manifold contraction is strong, the 2nd coordinate connecting $s(x, t) \rightarrow s$ can be neglected. We saw in example 14.3 how this

works. You might, by nature and temperament, take the dark view: Rössler has helpful properties, namely insanely strong contraction along a 1-dimensional stable direction, that are not present in real problems, such as turbulence in a plane Couette flow. Thus, the lessons of chapter 14 are of no use when it comes to real plumbing. For this reason, both of the training examples to come, the billiards and the Hénon map, are of the Hamiltonian, phase-space preserving type and are as far from being insanely contracting as possible. Yet, to a thoughtful reader, they unfold themselves as pages of a book.

Assign to each d -dimensional point $\hat{x} \in L_q$ a coordinate $s = s(\hat{x})$, whose value is the Euclidean arclength (15.4), to x_q measured along the 1-dimensional $\mathcal{P}_q$ section of the x_q unstable manifold. Next, for a nearby point $\hat{x}_0 \notin L_q$ determine the point $\hat{x}_1 \in L_q$ which minimizes the Euclidean distance $(\hat{x}_0 - \hat{x}_1)^2$, and assign arc length coordinate value $s_0 = s(\hat{x}_1)$ to $\hat{x}_0$. In this way, an approximate 1-dimensional intrinsic coordinate system is built along the unstable manifold. This parametrization is useful if the non-wandering set is sufficiently thin that its perpendicular extent can be neglected, with every point on the non-wandering set assigned the nearest point on the base segment L_q .

Armed with this intrinsic curvilinear coordinate parametrization, we are now in a position to construct a 1-dimensional model of the dynamics on the non-wandering set. If $\hat{x}_n$ is the n th Poincaré section of a trajectory in neighborhood of x_q , and s_n is the corresponding curvilinear coordinate, then $s_{n+1} = f^{t_n}(s_n)$ models the full state space dynamics $\hat{x}_n \rightarrow \hat{x}_{n+1}$. We approximate $f(s_n)$ by a smooth, continuous 1-dimensional map $f : L_q \rightarrow L_q$ by taking $\hat{x}_n \in L_q$ and assigning to $\hat{x}_{n+1}$ the nearest base segment point $s_{n+1} = s(\hat{x}_{n+1})$.

15.2 Horseshoes

If you find yourself mystified by Smale's article abstract quoted on page 295, about 'the action (differentiable) of a Lie group G on a manifold M ', time has come to bring Smale to everyman. If you still remain mystified by the end of this chapter, reading chapter 19 might help. For example, the Liouville operators form a Lie group of symplectic or canonical transformations acting on the (q, p) manifold.

If a flow is locally unstable but globally bounded, any open ball of initial points will be stretched out and then folded. An example is a 3-dimensional invertible flow sketched in figure 14.7 (a) which returns a Poincaré section of the flow folded into a 'horseshoe' (we shall belabor this in figure 15.4). We now offer two examples of locally unstable but globally bounded flows, which return an initial area stretched and folded into a 'horseshoe', such that the initial area is intersected at most twice. We shall refer to such mappings with at most 2^n transverse self-intersections at the n th iteration as *once-folding* maps.

exercise 15.1

The first example is the 3-disk game of pinball figure 14.5, which, for sufficiently separated disks (see figure 14.6), is an example of a complete Smale

Figure 15.3: The 3-disk game of pinball of figure 14.5, generated by starting from disk 1, preceded by disk 2, coded in binary, as in figure 1.8 (e). (a) Strips $\mathcal{M}_{s,i,j}$ which have survived a bounce in the past and will survive a bounce in the future. (b) Iteration corresponds to the decimal point shift; for example, all points in the rectangle $[1.01]$ map into the rectangles $[0.10]$, $[0.11]$ in one iteration.

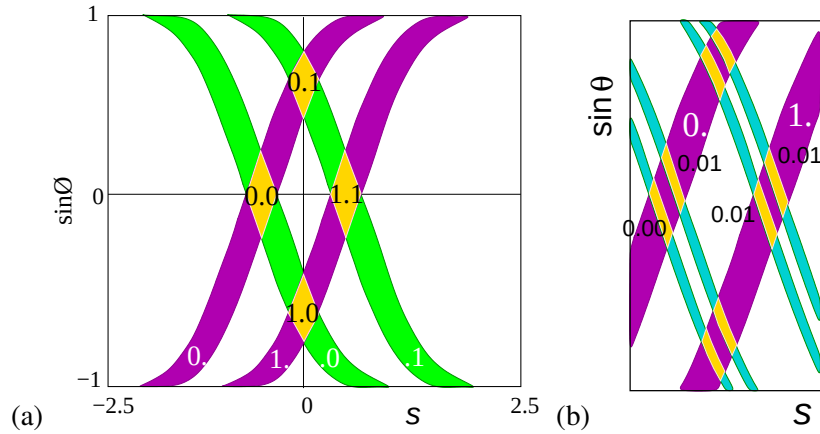
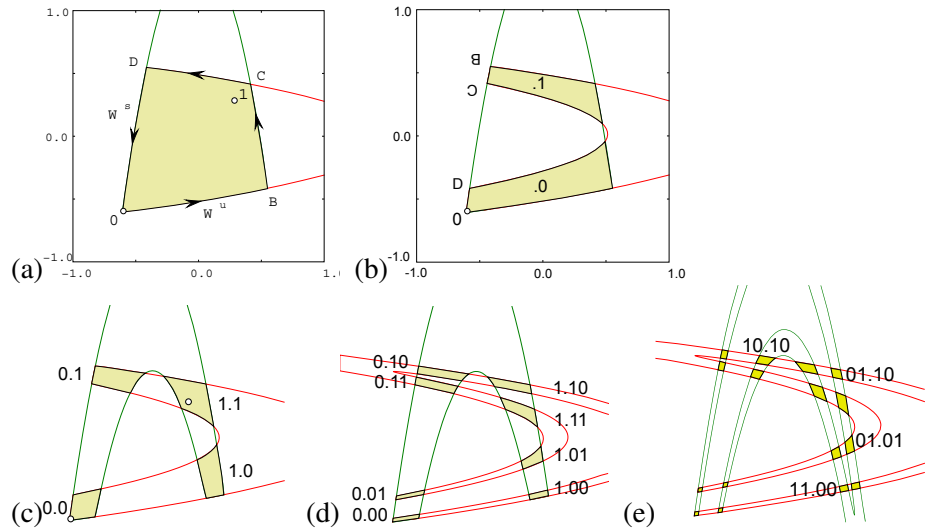


Figure 15.4: The Hénon map (15.20) for $a = 6$, $b = -1$: fixed point $\bar{0}$ with segments of its stable, unstable manifolds W^s , W^u , and fixed point $\bar{1}$. (a) Their intersection bounds the region $\mathcal{M} = OBCD$ which contains the non-wandering set Ω . (b) The intersection of the forward image $f(\mathcal{M})$ with $\mathcal{M}$, consists of two (future) strips $\mathcal{M}_0$, $\mathcal{M}_1$, with points BCD brought closer to fixed point $\bar{0}$ by the stable manifold contraction. (c) The intersection of the forward image $f(\mathcal{M})$ with the backward image $f^{-1}(\mathcal{M})$ is a four-region cover of Ω . (d) The intersection of the twice-folded forward horseshoe $f^2(\mathcal{M})$ is combined with the backward horseshoe $f^{-1}(\mathcal{M})$. (e) The intersection of $f^2(\mathcal{M})$ with $f^{-2}(\mathcal{M})$ is a 16-region cover of Ω . Iteration yields the complete Smale horseshoe non-wandering set Ω , i.e., the union of all non-wandering points of f , with every forward fold intersecting every backward fold. (P. Cvitanović and Y. Matsuoka)



horseshoe. We start by exploiting its symmetry to simplify it, and then partition its state space by its stable / unstable manifolds.



example 15.3
p. 315

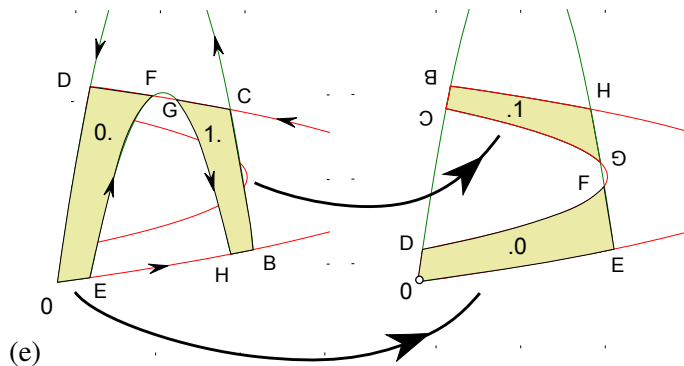
The 3-disk repeller does not really look like a ‘horseshoe;’ the ‘fold’ is cut out of the picture by allowing the pinballs that fly between the disks to fall off the table and escape. Next example captures the ‘stretch & fold’ horseshoe dynamics of return maps such as Rössler’s, figure 3.3.



example 15.4
p. 315

What is the significance of the subscript such as $_{011}$, which labels the $\mathcal{M}_{011}$ future strip? The two strips $\mathcal{M}_0, \mathcal{M}_1$ partition the state space into two regions labeled by the two-letter alphabet $\mathcal{A} = \{0, 1\}$. $S^+ = .011$ is the *future itinerary* for all $x \in \mathcal{M}_{011}$. Likewise, for the past strips all $x \in \mathcal{M}_{s-m \dots s-1 s_0}$ have the *past*

Figure 15.5: The dynamics maps two (past) strips $\mathcal{M}_0, \mathcal{M}_1$ into two (future) strips $\mathcal{M}_0, \mathcal{M}_1$. The corners are labeled to aid visualization. Note that the $BCGH$ strip is rotated by 180 degrees. (P. Cvitanović and Y. Matsuoka)



itinerary $S^- = s_{-m} \cdots s_{-1} s_0$. The partition we choose to pictorially present the regions that do not escape in m iterations is a matter of taste, as the backward strips are the preimages of the forward ones

$$\mathcal{M}_0 = f(\mathcal{M}_0), \quad \mathcal{M}_1 = f(\mathcal{M}_1).$$

Ω , the non-wandering set (2.5) of $\mathcal{M}$, is the union of all points whose forward and backward trajectories remain trapped for all time, given by the intersections of all images and preimages of $\mathcal{M}$:

$$\Omega = \left\{ x \mid x \in \lim_{m, n \rightarrow \infty} f^m(\mathcal{M}) \cap f^{-n}(\mathcal{M}) \right\}. \quad (15.6)$$

Two important properties of the Smale horseshoe are its *complete binary symbolic dynamics* and its *structurally stable*.

For a *complete* Smale horseshoe, every forward fold $f^n(\mathcal{M})$ intersects transversally every backward fold $f^{-m}(\mathcal{M})$, so a unique bi-infinite binary sequence can be associated to every element of the non-wandering set. A point $x \in \Omega$ is labeled by the intersection of its past and future itineraries $S(x) = \cdots s_{-2} s_{-1} s_0 s_1 s_2 \cdots$, where $s_n = s$ if $f^n(x) \in \mathcal{M}_s$, $s \in \{0, 1\}$ and $n \in \mathbb{Z}$.

remark A1.1

The system is said to be *structurally stable* if all intersections of forward and backward iterates of $\mathcal{M}$ remain transverse for sufficiently small perturbations $f \rightarrow f + \delta$ of the flow. For example, this is the case for slight displacements of the disks in the pinball problem or sufficiently small variations of the Hénon map parameters a, b . While structural stability is exceedingly desirable, it is also exceedingly rare. About this, more later.

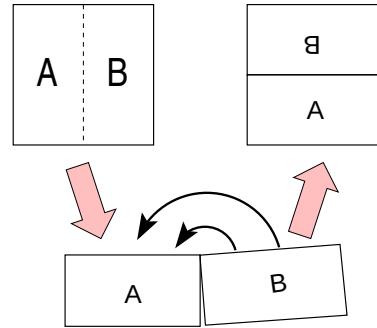
section 1.8

section 24.2

15.3 Symbol plane

Consider a system for which you have succeeded in constructing a covering symbolic dynamics, such as a well-separated 3-disk system. Now start moving the disks toward each other. At some critical separation, a disk will start blocking families of trajectories traversing the other two disks. The order in which trajectories disappear is determined by their relative ordering in space; the ones closest

Figure 15.6: The orientations involved with kneading a Danish pastry mimics the horseshoe dynamics of figure 15.5 by (1) squashing the unit square by factor 1/2, (2) stretching it by factor 2, and (3) folding the right half back over the left half.



to the intervening disk will be pruned first. Determining inadmissible itineraries requires that we relate the spatial ordering of trajectories to their time ordered itineraries.

exercise 15.8

So far we have rules that, given a state space partition, generate a *temporally* ordered itinerary for a given trajectory. Our next task is the converse: given a set of itineraries, what is the *spatial* ordering of corresponding points along the trajectories? In answering this question we will be aided by Smale's visualization of the relation between the topology of a flow and its symbolic dynamics by means of 'horseshoes', such as figure 15.4.

15.3.1 Kneading Danish pastry

The Danish pastry transformation, the simplest baker's transformation appropriate to Hénon type mappings, yields a binary coordinatization of all possible periodic points.

The symbolic dynamics of once-folding map is given by the Danish pastry transformation. This generates both the longitudinal and transverse alternating binary tree. The longitudinal coordinate is given by the head of a symbolic sequence; the transverse coordinate is given by the tail of the symbolic sequence. The dynamics on this space is given by symbol shift permutations: volume preserving with 2 expansion and 1/2 contraction.

For a better visualization of 2-dimensional non-wandering sets, fatten the intersection regions until they completely cover a unit square, as in figure 15.7.

We shall refer to such a 'map' of the topology of a given 'stretch & fold' dynamical system as the *symbol square*. The symbol square is a topologically accurate representation of the non-wandering set and serves as a street map for labeling its pieces. Finite memory of m steps and finite foresight of n steps partitions the symbol square into *rectangles* $[s_{-m+1} \cdots s_0.s_1s_2 \cdots s_n]$, such as those of figure 15.5. In the binary dynamics symbol square the size of such rectangle is $2^{-m} \times 2^{-n}$; it corresponds to a region of the dynamical state space, which contains all points that share common n future and m past symbols. This region maps in a nontrivial way in state space, but in the symbol square its dynamics is exceedingly simple. All of its points are mapped by a decimal point shift (14.12)

exercise 15.3
exercise 15.4

$$\sigma(\cdots s_{-2}s_{-1}s_0.s_1s_2s_3 \cdots) = \cdots s_{-2}s_{-1}s_0s_1.s_2s_3 \cdots, \quad (15.7)$$

$\delta = \delta(S^-)$ by the algorithm:

$$w_{n-1} = \begin{cases} w_n & \text{if } s_n = 0 \\ 1 - w_n & \text{if } s_n = 1 \end{cases}, \quad w_0 = s_0$$

$$\delta(S^-) = 0.w_0w_{-1}w_{-2}\dots = \sum_{n=1}^{\infty} w_{1-n}/2^n. \quad (15.8)$$

Such formulas are best derived by solitary contemplation of the action of a folding map, in the same way we derived the future topological coordinate (14.4).

The coordinate pair (δ, γ) associates a point (x, y) in the state space Cantor set of figure 15.4 to a point in the symbol square of figure 15.9, preserving the topological ordering. The symbol square $[\delta, \gamma]$ serves as a topologically faithful representation of the non-wandering set of any once-folding map, and this aids us in partitioning the set and ordering the partitions for any flow of this type.



fast track:
chapter 16, p. 321

15.4 Prune Danish

Anyone know where I can get a good prune Danish in Charlotte? I mean a real NY Jewish bakery kind of prune Danish!

— Googled



In general, not all possible symbol sequences are realized as physical trajectories. Trying to get from ‘here’ to ‘there’, we might find that a short path is excluded by some obstacle, such as a disk that blocks the path or a mountain. In order to enumerate orbits correctly, we need to *prune* the inadmissible symbol sequences, i.e., describe the grammar of the admissible itineraries.

The complete Smale horseshoe dynamics discussed so far is rather straightforward and sets the stage for situations that more closely resemble real life. A generic once-folding map does not yield a complete horseshoe; some of the horseshoe pieces might be *pruned*, i.e., not realized for particular parameter values of the mapping. In 1 dimension, the criterion for whether a given symbolic sequence is realized by a given unimodal map is easily formulated. Any orbit that strays to the right of the value computable from the *kneading sequence* (the orbit of the critical point (14.5)) is pruned. This is a topological statement, independent of a particular unimodal map. Our objective is to generalize this notion to 2-dimensional once-folding maps.

Adjust the parameters of a once-folding map so that the intersection of the backward and forward folds is still transverse, but no longer complete, as in figure 15.10. The utility of the symbol square lies in the fact that the surviving,

Figure 15.10: (left) An incomplete Smale horseshoe: the inner forward fold does not intersect the outer backward fold. (right) The primary pruned region in the symbol square and the corresponding forbidden binary blocks.

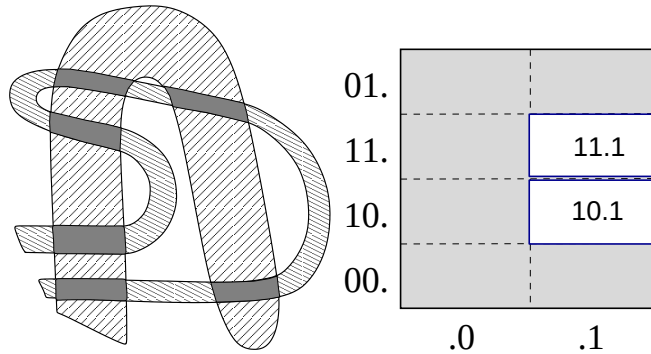
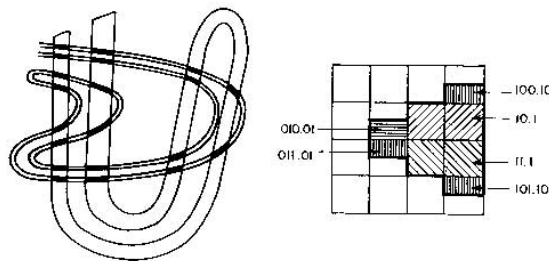


Figure 15.11: (a) An incomplete Smale horseshoe which illustrates (b) the monotonicity of the pruning front: the thick line which delineates the left border of the primary pruned region is monotone on each half of the symbol square. The backward folding in this figure and figure 15.10 is schematic - in invertible mappings there are further missing intersections, all obtained by the forward and backward iterations of the primary pruned region.



admissible itineraries still maintain the same relative spatial ordering as for the complete case.

In the example of figure 15.10, the rectangles $[10.1]$, $[11.1]$ have been pruned, and consequently *any* itinerary containing substrings $b_1 = 101$, $b_2 = 111$ are inadmissible or *pruned*. The symbol dynamics is a subshift of finite type (14.17).

We refer to the left border of this primary pruned region as the *pruning front*, and another example of a pruning front is drawn in figure 15.11 (b). We call it a ‘front’ as it can be visualized as a border between admissible and inadmissible; any trajectory whose points would fall to the right of the front in figure 15.11 is inadmissible, i.e., pruned. The pruning front is a complete description of the symbolic dynamics of once-folding maps (read sect. 15.4.1). The pruning front is a 2-dimensional generalization of the 1-dimensional kneading sequence (14.5). The location of each vertical step in the pruning front is the kneading sequence of the corresponding primary turnback of the unstable manifold.

In the example of figure 15.10, there are a finite number of grammar rules and a total of two forbidden blocks 101, 111. For now we concentrate on this kind of pruning (‘subshifts of finite type’), because it is particularly clean and simple.

15.4.1 Pruning front conjecture

No matter how far down the wrong road you’ve gone, turn back.

— Turkish proverb



The pruning front conjecture offers a complete description of the symbolic dy-

namics of once-folding maps in the same sense in which the kneading sequence defines the symbolic dynamics of a 1-dimensional unimodal map. The intuition behind this conjecture is that the folding induced by a single iteration is the primary folding, and all other folds (turnbacks, homoclinic tangencies) are images or preimages of the primary ones. The topology puts two constraints on the form of a pruning front for once-folding maps:

1. The pruning front is symmetric across the horizontal $1/2$ line.
2. The pruning front is *monotone* across either half of the symbol square.

This is a consequence of the deterministic foliation; inner folds cannot pierce through the outer folds and, therefore, have the same number or fewer transverse sections than the outer ones.

Our strategy is the following: we first construct the symbol square, the 2-dimensional ‘NYC subway map’ of the topology of a given ‘stretch & fold’ dynamical system, as illustrated in figure 15.7. The symbol square is a ‘road map’ in which the various sheets of the stable and unstable manifolds are represented by straight sections, and the topology is preserved. The nearby periodic points in the symbol square represent nearby periodic points in the state space. Next we separate the admissible and the forbidden motions by means of a ‘pruning front’, a boundary between the two kinds of orbits. This approach assumes the following:

- (i) *The partition conjecture*: the non-wandering set of a once-folding map can be described by a subset of a complete Smale horseshoe, partitioned by the set of primary turning points.
- (ii) *The pruning-front conjecture*: kneading values of the set of all primary turning points separate the admissible from the forbidden orbits, and there are no other pruning rules.
- (iii) *Multimodal map approximation*: A 2-dimensional once-folding map can be systematically approximated by a sequence of 1-dimensional n -folding maps.

The intuition behind these conjectures is that the folding induced by a single iteration is the primary folding, and all other folds (turning points, homoclinic tangencies) are images or preimages of the primary ones. The asymptotic object is a collection of infinitely many 1-dimensional sheets, and the pruning front is the set of the corresponding kneading sequences (14.5), one for each 1-dimensional sheet.



fast track:
chapter 16, p. 321

Though a useful tool, Markov partitioning is not without drawbacks. One glaring shortcoming is that Markov partitions are not unique. Any of many different

Table 15.1: Correspondence between the Z_2 symmetry reduced cycles $\tilde{p}$ and the full state space periodic orbits p , together with their multiplicities m_p . Also listed are the two shortest cycles (length 6) related by time reversal, but distinct under D_1 .

$\tilde{p}$	p	m_p
1	+	2
0	-+	1
01	-- ++	1
001	-++	2
011	--- +++	1
0001	-+-- +-++	1
0011	-+++	2
0111	---- +++++	1
00001	-+-+ -	2
00011	-+--- +-+++	1
00101	-++-- +- -++	1
00111	-+--- +-+++	1
01011	--+++	2
01111	----- ++++++	1
001011	-++-- - +- -++	1
001101	-+++ - - +- - -++	1

partitions might do the job. The Z_2 - and D_3 -equivariant systems that we discuss next offer a simple illustration of different Markov partitioning strategies for the same dynamical system.

15.5 Recoding, symmetries, tilings



In chapter 11 we made a claim that if there is a symmetry of dynamics, we must use it. Here we shall show how to use it, on two concrete examples, and in chapter 25 we shall be handsomely rewarded for our labors. First, the simplest example of equivariance, a single ‘reflection’ $D_1 = Z_2$ group of example 11.4.



example 15.6
p. 316

Next, let us take the old pinball game and ‘quotient’ the state space by the symmetry, or ‘desymmetrize.’ As the three disks are equidistantly spaced, our game of pinball has a sixfold symmetry. For instance, the cycles $\overline{12}$, $\overline{23}$, and $\overline{13}$ in figure 15.12 are related to each other by rotation by $\pm 2\pi/3$ or, equivalently, by a relabeling of the disks. We exploit this symmetry by recoding, as in (15.19).



example 15.5
p. 316

exercise 14.1
exercise 15.7
exercise 15.8
exercise 17.2

Binary symbolic dynamics has two immediate advantages over the ternary one; the prohibition of self-bounces is automatic, and the coding utilizes the symmetry of the 3-disk pinball game in an elegant manner.

The 3-disk game of pinball is tiled by six copies of the *fundamental domain*, a one-sixth slice of the full 3-disk system, with the symmetry axes acting as reflect-



exercise 14.2

Figure 15.12: The 3-disk game of pinball with the disk radius : center separation ratio $a:R = 1:2.5$. (a) 2-cycles $\overline{12}$, $\overline{13}$, $\overline{23}$, and 3-cycles $\overline{123}$ and $\overline{132}$ (not drawn). (b) The fundamental domain, i.e., the small $1/6$ th wedge indicated in (a), consisting of a section of a disk, two segments of symmetry axes acting as straight mirror walls, and an escape gap. The above five cycles restricted to the fundamental domain are the two fixed points $\overline{0}$, $\overline{1}$. See figure 11.1 for cycle $\overline{10}$ and further examples.

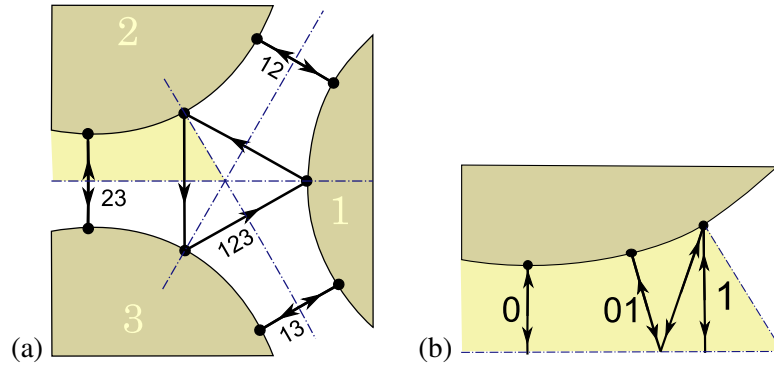


Table 15.2: D_3 correspondence between the binary labeled fundamental domain prime cycles $\tilde{p}$ and the full 3-disk ternary labeled cycles p , together with the D_3 transformation that maps the end point of the $\tilde{p}$ cycle into the irreducible segment of the p cycle, see example 11.5. White spaces in the above ternary sequences mark repeats of the irreducible segment; for example, the full space 12-cycle 1212 3131 2323 consists of 1212 and its symmetry related segments 3131, 2323. The multiplicity of p cycle is $m_p = 6n_{\tilde{p}}/n_p$. The shortest pair of fundamental domain cycles related by time reversal (but no spatial symmetry) are the 6-cycles 001011 and 001101.

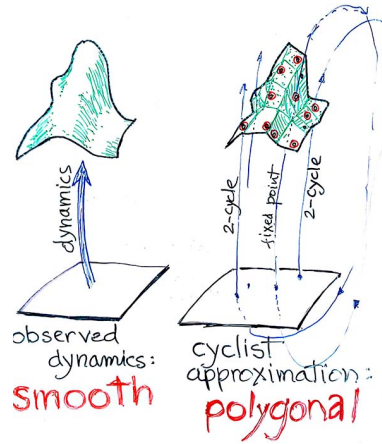
$\tilde{p}$	p	$\mathbf{g}_{\tilde{p}}$	$\tilde{p}$	p	$\mathbf{g}_{\tilde{p}}$
0	1 2	σ_{12}	000001	121212 131313	σ_{23}
1	1 2 3	C	000011	121212 313131 232323	C^2
01	12 13	σ_{23}	000101	121213	e
001	121 232 313	C	000111	121213 212123	σ_{12}
011	121 323	σ_{13}	001011	121232 131323	σ_{23}
0001	1212 1313	σ_{23}	001101	121231 323213	σ_{13}
0011	1212 3131 2323	C^2	001111	121231 232312 313123	C
0111	1213 2123	σ_{12}	010111	121312 313231 232123	C^2
00001	12121 23232 31313	C	011111	121321 323123	σ_{13}
00011	12121 32323	σ_{13}	0000001	1212121 2323232 3131313	C
00101	12123 21213	σ_{12}	0000011	1212121 3232323	σ_{13}
00111	12123	e	0000101	1212123 2121213	σ_{12}
01011	12131 23212 31323	C	0000111	1212123	e
01111	12132 13123	σ_{23}	...	...	...

ing mirrors, see figure 15.12(b). Every global 3-disk trajectory has a corresponding fundamental domain mirror trajectory obtained by replacing every crossing of a symmetry axis by a reflection. Depending on the symmetry of the full state space trajectory, a repeating binary alphabet block corresponds either to the full periodic orbit or to a relative periodic orbit (examples are shown in figure 15.12 and table 15.2). A relative periodic orbit corresponds to a periodic orbit in the fundamental domain.

Table 15.2 lists some of the shortest binary periodic orbits, together with the corresponding full 3-disk symbol sequences and orbit symmetries. For a number of deep reasons that will be elucidated in chapter 25, life is much simpler in the fundamental domain than in the full system, so whenever possible our computations will be carried out in the fundamental domain.

 example 15.7
p. 317

Figure 15.13: Reduction of a continuous-time flow (left frame) to a set of return maps (right frame), with a point on 1-cycle and the two cycle points of a 2-cycle used as template points.



15.6 Charting the state space

In simple examples, such as the Rössler example 3.2, a single Poincaré section suffices, but this is rarely the case for flows of physical interest. In this section (skip it on first reading) we commence a discussion of the general case.



A Poincaré section is constructed by picking a ‘template’ point $\hat{x}'$ within a state space region of interest and defining a hypersurface (3.2) that goes through the template point. In theory, this Poincaré section could be any $(d-1)$ -dimensional manifold. In practice, a hyperplane (3.5) is the most convenient. Its a natural choice for the vector normal to the section being $\hat{n} = v(\hat{x}')$, the velocity field at the template point $\hat{x}'$. This Poincaré section $\hat{x} \in \mathcal{P}$ is a hyperplane,

$$v' \cdot (\hat{x} - \hat{x}') = 0, \quad v' = v(\hat{x}'), \quad (15.9)$$

appendix 7.2

normal to the flow direction v' at the template point $\hat{x}'$. Such section cuts the nearby trajectories transversally, and is a good description of solutions similar to the given template.

So one hyperspace $\mathcal{P}$ will, in general, not suffice. A more insightful picture of the dynamics is obtained by partitioning the state space into N qualitatively distinct regions $\{\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_N\}$ and constructing a Poincaré section per region, global *atlas* of the state space composed of N local Poincaré sections $\mathcal{P}^{(j)}$ or *charts*, each one capturing a neighborhood of a qualitatively prominent state $\hat{x}^{(j)} \in S$. We shall refer to these states as *templates*, each represented in the state space $\mathcal{M}$ of the system by a *template point* $\{\hat{x}^{(1)}, \hat{x}^{(2)}, \dots, \hat{x}^{(N)}\}$.

section 14.1

Our Poincaré section is a hyperplane. If we pick another template point $\hat{x}^{(2)}$, it comes along with its own section hyperplane. The $(d-1)$ -dimensional Poincaré sections for an adjacent pair of template intersects in a ‘ridge’ (‘boundary’, ‘edge’), a $(d-2)$ -dimensional hyperplane, easy to compute. Follow an ant (the sequence of return map iterates) as it progresses along the Poincaré section $\mathcal{P}^{(1)}$. The moment $(\hat{x}^{(1)}(\tau) - \hat{x}^{(2)}) \cdot \hat{n}^{(2)}$ changes sign, the ant has crossed the ridge, we switch the Poincaré section, and the ant continues its merry stroll now confined to the $\mathcal{P}^{(2)}$ section. Each Poincaré section $\mathcal{P}^{(j)}$ provides a local chart at

$\hat{x}'^{(j)}$ for a neighborhood of an important, qualitatively distinct class of solutions. Together they ‘Voronoi’ tessellate the curved manifold in which the reduced dynamics is replaced by a finite set of mappings between hyperplane tiles. An example is the periodic-orbit implementation of the idea of state space tessellation by neighborhoods of recurrent points, so dear to professional cyclists, illustrated in figure 15.13.

For a given dynamical flow, the physical task is to pick a minimal set of qualitatively distinct templates. The state space might be filled by all kinds of highly unstable, never revisited equilibria and relative periodic orbits. The choice of templates should reflect the dynamically prominent states seen in the long-time simulations of system dynamics. We have only vague advice on how to pick a single Poincaré section (see sect. 3.1.2), and we have nearly no advice on how to systematically pick a set of ‘good’ templates. The only guidance we can offer here is that associated section tiles should be as large as possible, but still sufficiently small enough to exclude orbit tangencies, i.e., stop before crossing their section borders (3.8). Ideally, one would like to pick as few templates as possible in figure 15.13. Once templates are picked, the rest is geometry of hyperplanes, so checking whether the section border is on the far side of the tile edge (ridge between two sections) is a fast, linear computation.

There is a rub, though - you need to know how to pick the neighboring templates. Perhaps a glance at figure 15.13 helps visualize the problem. Imagine that the tiles belong to the Poincaré sections through template points on these orbits. One could slide templates along their trajectories until the pairs of straight line segments connecting neighboring template points are minimized, but that seems a bit arbitrary. At this time we have no advice as how to ‘synchronize’ the templates relative to each other. The astute reader will instantly recognize this as the problem of ‘local gauge invariance’ or ‘gauge fixing’ of Quantum Field Theory and General Relativity.

15.6.1 Navigating the Poincaré-charted state space

Our goal now is to replace the continuous-time dynamics by a set of return maps between a set of hyperplane sections, as in figure 15.13. The flat hyperplane (3.5) is an *ad hoc* construct; one Poincaré section rarely suffices to capture all of the dynamics of interest. Instead we chart the state space by partitioning it into N qualitatively distinct regions $\{\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_N\}$. Successive trajectory intersections, along with the set of $(d-1)$ -dimensional hypersurfaces $\mathcal{P}_s$ embedded in the d -dimensional state space $\mathcal{M}$, define the set of $(d-1) \rightarrow (d-1)$ return maps

section 14.1

$$\begin{aligned} \hat{x}_{n+1} &= P_{s_{n+1}s_n}(\hat{x}_n) = f^{\tau(\hat{x}_n)}(\hat{x}_n) \\ \hat{x}_{n+1} &\in \mathcal{P}^{s_{n+1}}, \quad \hat{x}_n \in \mathcal{P}^{s_n}, \quad s \in \{1, 2, \dots, N\}. \end{aligned} \quad (15.10)$$

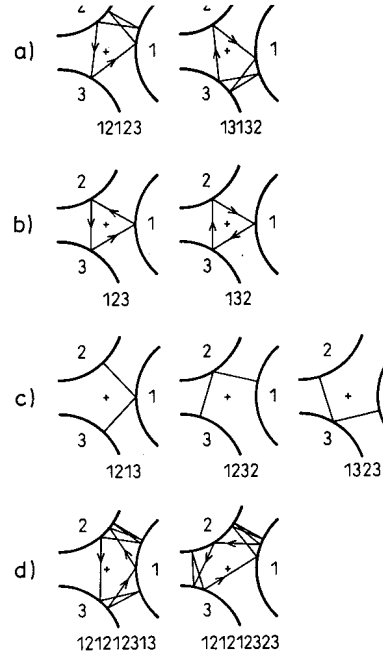


Figure 15.14: Some examples of 3-disk cycles: (a) $\overline{12123}$ and $\overline{13132}$ are mapped into each other by the flip across 1 axis. Similarly, (b) $\overline{123}$ and $\overline{132}$ are related by flips, and (c) $\overline{1213}$, $\overline{1232}$ and $\overline{1323}$ by rotations. (d) The cycles $\overline{121212313}$ and $\overline{121212323}$ are related by rotation *and* time reversal. These symmetries are discussed in chapter 11 (from ref. [15]).

The d -dimensional continuous time flow is thus reduced to discrete time composition

$$P_{s_0 s_1 \cdots s_n} = P_{s_n s_{n-1}} \circ \cdots \circ P_{s_2 s_1} \circ P_{s_1 s_0}$$

of a set of return maps (15.10) that map the coordinates of Poincaré section $\mathcal{P}_{s_n}$ to those of $\mathcal{P}_{s_{n+1}}$, the next section traversed by a given trajectory.

If a trajectory traverses regions $\mathcal{M}_{s_0} \rightarrow \mathcal{M}_{s_1} \rightarrow \cdots \rightarrow \mathcal{M}_{s_n}$, the sequence $s_0 s_1 \cdots s_n = s_n \leftarrow \cdots \leftarrow s_1 \leftarrow s_0$ is said to be *admissible*. The *return map* P_{s_0} from section $\mathcal{P}_{s_0}$ to itself has a contribution from any admissible returning (periodic, $s_n = s_0$) sequence of compositions

section 14.6

$$P_{s_0 s_1 \cdots s_{n-1} s_0} = P_{s_0 s_{n-1}} \circ \cdots \circ P_{s_2 s_1} \circ P_{s_1 s_0}. \quad (15.11)$$

The next example offers an unambiguous set of such Poincaré sections which do double duty, providing us both with an exact representation of dynamics in terms of maps and with symbolic dynamics, a subject that we will return to in chapter 14.

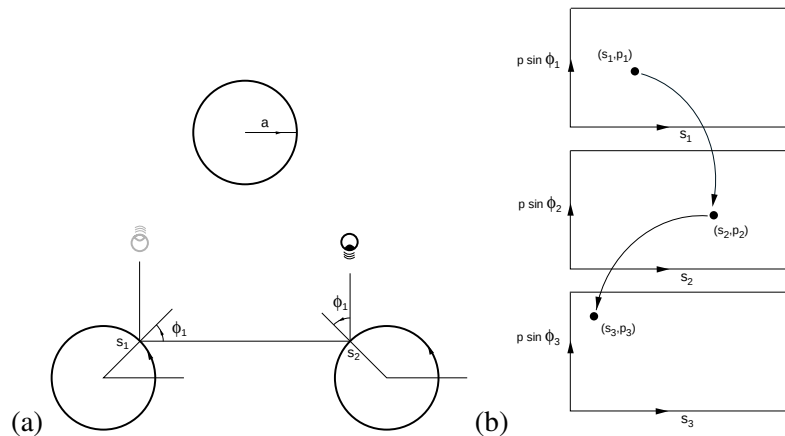
chapter 14



example 15.8
p. 317

Billiard dynamics is exceptionally simple - free flight segments, followed by specular reflections at boundaries. Billiard boundaries are the obvious choice for Poincaré sections. For a general flow one is never so lucky. Also, so far we have discussed only flows with one continuous parameter (time). The general case of N -parameter continuous symmetries is discussed in chapter 12.

Figure 15.15: (a) Poincaré section coordinates for the 3-disk game of pinball. (b) Collision sequence $(s_1, p_1) \mapsto (s_2, p_2) \mapsto (s_3, p_3)$ from the boundary of a disk to the boundary of the next disk presented in the Poincaré section coordinates and coded by the return maps sequence $P_{3 \leftarrow 2} P_{2 \leftarrow 1}$.



Résumé

In the preceding and this chapter, we start with a d -dimensional state space and end with a 1-dimensional return map description of the dynamics. The arc-length parametrization of the unstable manifold maintains the 1-to-1 relation of the *full* d -dimensional state space dynamics and its 1-dimensional return-map representation. To high accuracy *no information about the flow is lost* by its 1-dimensional return map description. We explain why Lorenz equilibria are connected to heteroclinic orbits (it is not due to the symmetry) and how to generate all periodic orbits of Lorenz flow up to a given length. This we do, in contrast to the rest of the thesis, without any group-theoretical jargon to blind you with.

For 1-dimensional maps the folding point is the critical point, and easy to determine. In higher dimensions, the situation is not so clear - one can attempt to determine the (fractal set of) folding points by looking at their higher iterates. Due to the contraction along stable manifolds, the fold gets exponentially sharper at each iterate. In practice this set is essentially uncontrollable for the same reason the flow itself is chaotic - exponential growth of errors. We prefer to determine a folding point by bracketing it by longer and longer cycles which can be determined accurately using variational methods of chapter 34, irrespective of their period.

For a generic dynamical system a subshift of finite type is the exception rather than the rule. Its symbolic dynamics can be arbitrarily complex; even for the logistic map the grammar is finite only for special parameter values. Only some repelling sets (like our game of pinball) and a few purely mathematical constructs (called Anosov flows) are structurally stable. For most systems of interest, an infinitesimal perturbation of the flow destroys and/or creates an infinity of trajectories, and specification of the grammar requires determination of pruning blocks of arbitrary length. The repercussions are dramatic and counterintuitive; for example, the transport coefficients such as the deterministic diffusion constant of sect. 24.2 are emphatically *not* smooth functions of the system parameters. The importance of symbolic dynamics is often under appreciated. As we shall see in chapters 23 and 28, the existence of a finite grammar is the crucial prerequisite for construction of zeta functions with nice analytical properties. This generic lack of structural stability is what makes nonlinear dynamics so hard.

section 24.2

The conceptually simpler finite subshift Smale horseshoes suffice to motivate most of the key concepts that we shall need for the time being. Our strategy is akin to bounding a real number by a sequence of rational approximants; we converge toward the non-wandering set under investigation by a sequence of self-similar Cantor sets. The rule that everything to one side of the pruning front is forbidden is striking in its simplicity. The pruning front cleanly cuts out a *compact* region in the symbol square, and that is all - there are no additional pruning rules. Alternatively, pruning a Cantor set embedded within some larger Cantor set is more cumbersome. A ‘self-similar’ Cantor set (in the sense in which we use the word here) is a Cantor set equipped with a *subshift of finite type* symbol dynamics, i.e., the corresponding grammar can be stated as a finite number of pruning rules, each forbidding a finite subsequence $\_s_1 s_2 \dots s_n \_$. Here the notation $\_s_1 s_2 \dots s_n \_$ stands for n consecutive symbols $s_1, s_2, \dots, s_n$, preceded and followed by arbitrary symbol strings.

The symbol square is a useful tool in transforming topological pruning into pruning rules for inadmissible sequences. Those are implemented by constructing transition matrices and/or graphs, see chapters 17 and 18.

Commentary

Remark 15.1. Stable/unstable manifolds. For pretty hand-drawn pictures of invariant manifolds, see Abraham and Shaw [1]. Construction of invariant manifolds by map iteration is described in Simo [39]. Fixed point stable / unstable manifolds and their homoclinic and heteroclinic intersections can be computed using DsTool [3, 16, 32]. Unstable manifold turning points were utilized in refs. [8–11, 20, 38] to partition state space and prune inadmissible symbol sequences. The arclength parameterized return maps were introduced by Christiansen *et al.* [7], and utilized in ref. [33]. Even though no dynamical system has been studied more exhaustively than the Lorenz equations, the analysis of sect. 14.2 is new. The desymmetrization follows Gilmore and Lettelier [18], but the key new idea is taken from Christiansen *et al.* [7]: the arc-length parametrization of the unstable manifold maintains the 1-to-1 relation of the *full* d -dimensional state space dynamics and its 1-dimensional return-map representation, in contrast to 1-dimensional *projections* of the $(d-1)$ -dimensional Poincaré section return maps previously deployed in the literature. In other words, to high accuracy *no information about the flow is lost* by its 1-dimensional return map description.

Remark 15.2. Smale horseshoe. S. Smale understood clearly that the crucial ingredient in the description of a chaotic flow is the topology of its non-wandering set, and he provided us with the simplest visualization of such sets as intersections of Smale horseshoes. In retrospect, much of the material covered here can already be found in Smale’s fundamental paper [41], but an engineer or a scientist who has run into a chaotic time series in his laboratory might not know that he is investigating the action (differentiable) of a Lie group G on a manifold M , and that the Lefschetz trace formula is the way to go.

We have tried to explain the geometric picture the best we could in the static text format, but there is no substitute for dynamics but the dynamics itself. We found Demidov’s “Chaotic maps” [14] simulations of the Hénon map particularly helpful in explaining how horseshoes partition the non-wandering sets. For a detailed discussion of the Hamiltonian, area-preserving Hénon map (sketched in example 15.9), see Li and Tomsovic [34].

Symbolic dynamics of heteroclinic and homoclinic connections is discussed by Hagiwara and Shudo [24].

section 14.6

Remark 15.3. Pruning fronts. The ‘partition conjecture’ is due to Grassberger and Kantz [21]. The notion of a pruning front and the ‘pruning-front conjecture’ was formulated by Cvitanović *et al.* [11], and it was developed by K.T. Hansen for a number of dynamical systems in Hansen’s Ph.D. thesis [26] and a series of papers [25]–[27]. The ‘multimodal map approximation’ is described in Hansen’s thesis [26]. The thesis remains the most accessible exposition of the pruning theory and its applications. Detailed studies of pruning fronts are carried out in refs. [12, 13, 19]; ref. [22] is the most detailed study carried out so far. The rigorous theory of pruning fronts has been developed by Y. Ishii [29–31] for the Lozi map by and A. de Carvalho [5, 6] for a very general setting. Beyond the orbit pruning and its infinity of admissible unstable orbits, an attractor of Hénon type may also own an infinity of attractive orbits coexisting with the strange attractor [35, 36]. We offer heuristic arguments and numerical evidence that the coexistence of attractive orbits does not destroy the strange attractor/repeller, which is also in this case described by the 2-dimensional Danish pastry plot.

Remark 15.4. Cat map symbolic dynamics. (Continued from remark 8.3) Adler [2] *Symbolic dynamics and Markov partitions*, is an excellent overview of symbolic dynamics techniques. Adler writes: “The decimal expansion of real numbers, familiar to us all, has a dramatic generalization to representation of dynamical system orbits by symbolic sequences. The natural way to associate a symbolic sequence with an orbit is to track its history through a partition. But in order to get useful symbolism, one needs to construct a partition with special properties. In this work we develop a general theory of representing dynamical systems by symbolic systems by means of so-called Markov partitions. We apply the results to one of the more tractable examples: namely, hyperbolic automorphisms of the two dimensional torus.”

“The first use of infinite sequences of symbols to describe orbits is attributed to a nineteenth century work of Hadamard [23]. However the present work is rooted in something very much older and familiar to us all: namely, the representation of real numbers by infinite binary expansions. Markov partitions for hyperbolic automorphisms acting on the two dimensional torus have come to play a pervasive role in understanding the dynamics of general hyperbolic systems, e.g., Anosov diffeomorphisms, axiom A diffeomorphisms, and pseudo-Anosov diffeomorphisms. Sinai [40] constructed Markov partitions for Anosov diffeomorphisms (a simpler treatment can be found in Bowen [4]). This class of maps includes hyperbolic automorphisms of n -dimensional tori, $n > 2$. ” (continued in remark A24.1)

References

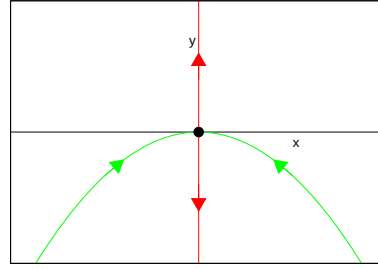
- [1] R. H. Abraham and C. D. Shaw, *Dynamics - The Geometry of Behavior* (Wesley, Reading, MA, 1992).
- [2] R. L. Adler, “Symbolic dynamics and Markov partitions”, *Bull. Amer. Math. Soc.* **35**, 1–56 (1998).
- [3] A. Back, J. Guckenheimer, M. R. Myers, F. J. Wicklin, and P. A. Worfolk, “DsTool: Computer assisted exploration of dynamical systems”, *Notices Amer. Math. Soc.* **39**, 303–309 (1992).

- [4] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms* (Springer, New York, 1975).
- [5] A. de Carvalho, “Pruning fronts and the formation of horseshoes”, *Ergod. Theor. Dynam. Syst.* **19**, 851–894 (1999).
- [6] A. de Carvalho and T. Hall, “How to prune a horseshoe”, *Nonlinearity* **15**, R19–R68 (2002).
- [7] F. Christiansen, P. Cvitanović, and V. Putkaradze, “Hopf’s last hope: Spatiotemporal chaos in terms of unstable recurrent patterns”, *Nonlinearity* **10**, 55–70 (1997).
- [8] F. Christiansen and A. Politi, “Generating partition for the standard map”, *Phys. Rev. E* **51**, R3811–R3814 (1995).
- [9] F. Christiansen and A. Politi, “Symbolic encoding in symplectic maps”, *Nonlinearity* **9**, 1623–1640 (1996).
- [10] F. Christiansen and A. Politi, “Guidelines for the construction of a generating partition in the standard map”, *Physica D* **109**, 32–41 (1997).
- [11] P. Cvitanović, G. H. Gunaratne, and I. Procaccia, “Topological and metric properties of Hénon-type strange attractors”, *Phys. Rev. A* **38**, 1503–1520 (1988).
- [12] G. D’Alessandro, P. Grassberger, S. Isola, and A. Politi, “On the topology of the Hénon map”, *J. Phys. A* **23**, 5285–5294 (1990).
- [13] G. D’Alessandro, S. Isola, and A. Politi, “Geometric properties of the pruning front”, *Progr. Theor. Phys.* **86**, 1149–1157 (1991).
- [14] E. Demidov, *Chaotic maps*, 2009.
- [15] B. Eckhardt, G. Russberg, P. Cvitanović, P. E. Rosenqvist, and P. Scherer, “Pinball scattering”, in *Quantum Chaos: Between Order and Disorder*, edited by G. Casati and B. Chirikov (Cambridge Univ. Press, Cambridge UK, 1995).
- [16] J. P. England, B. Krauskopf, and H. M. Osinga, “Computing one-dimensional stable and unstable sets of planar maps without the inverse”, *SIAM J. Appl. Dyn. Syst.* **3**, 161–190 (2004).
- [17] J. F. Gibson, J. Halcrow, and P. Cvitanović, “Visualizing the geometry of state-space in plane Couette flow”, *J. Fluid Mech.* **611**, 107–130 (2008).
- [18] R. Gilmore and C. Letellier, *The Symmetry of Chaos* (Oxford Univ. Press, Oxford, 2007).
- [19] F. Giovannini and A. Politi, “Homoclinic tangencies, generating partitions and curvature of invariant manifolds”, *J. Phys. A* **24**, 1837–1848 (1991).
- [20] P. Grassberger, R. Badii, and A. Politi, “Scaling laws for invariant measures on hyperbolic and nonhyperbolic attractors”, *J. Stat. Phys.* **51**, 135–178 (1988).
- [21] P. Grassberger and H. Kantz, “Generating partitions for the dissipative Hénon map”, *Phys. Lett. A* **113**, 235–238 (1985).
- [22] P. Grassberger, H. Kantz, and U. Moenig, “On the symbolic dynamics of Hénon map”, *J. Phys. A* **22**, 5217–5230 (1989).

- [23] J. S. Hadamard, “Les surfaces à courbures opposées et leurs lignes géodésique”, *J. Math. Pures Appl.* **4**, 27–73 (1898).
- [24] R. Hagiwara and A. Shudo, “Grammatical complexity for two-dimensional maps”, *J. Phys. A* **37**, 10545 (2004).
- [25] K. T. Hansen, “Pruning of orbits in four-disk and hyperbola billiards”, *Chaos* **2**, 71–75 (1992).
- [26] K. T. Hansen, *Symbolic Dynamics in Chaotic systems*, PhD thesis (Univ. of Oslo, Oslo, Norway, 1993).
- [27] K. T. Hansen and P. Cvitanović, “Bifurcation structures in maps of Hénon type”, *Nonlinearity* **11**, 1233–1261 (1998).
- [28] M. Hénon, “A two-dimensional mapping with a strange attractor”, *Commun. Math. Phys.* **50**, 69 (1976).
- [29] Y. Ishii, “Towards a kneading theory for Lozi mappings I: A solution of the pruning front conjecture and the first tangency problem”, *Nonlinearity* **10**, 731–747 (1997).
- [30] Y. Ishii, “Towards a kneading theory for Lozi mappings. II: Monotonicity of the topological entropy and Hausdorff dimension of attractors”, *Commun. Math. Phys.* **190**, 375–394 (1997).
- [31] Y. Ishii and D. Sands, “Monotonicity of the Lozi family near the tent-maps”, *Commun. Math. Phys.* **198**, 397–406 (1998).
- [32] B. Krauskopf and H. M. Osinga, “Investigating torus bifurcations in the forced Van der Pol oscillator”, in *Numerical Methods for Bifurcation Problems and Large-Scale Dynamical Systems*, edited by E. Doedel and L. Tuckerman (Springer, New York, 2000), pp. 199–208.
- [33] Y. Lan and P. Cvitanović, “Unstable recurrent patterns in Kuramoto-Sivashinsky dynamics”, *Phys. Rev. E* **78**, 026208 (2008).
- [34] J. Li and S. Tomsovic, “Exact relations between homoclinic and periodic orbit actions in chaotic systems”, *Phys. Rev. E* **97**, 022216 (2017).
- [35] S. E. Newhouse, “Diffeomorphisms with infinitely many sinks”, *Topology* **13**, 9–18 (1974).
- [36] S. E. Newhouse, “The abundance of wild hyperbolic sets and non-smooth stable sets for diffeomorphisms”, *Publ. Math. IHES* **50**, 101–152 (1979).
- [37] H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste* (Guthier-Villars, Paris, 1899).
- [38] A. Politi, R. Badii, and P. Grassberger, “On the geometric structure of non-hyperbolic attractors”, *J. Phys. A* **21**, L763–L769 (1988).
- [39] C. Simó, “On the analytical and numerical approximation of invariant manifolds”, in *Les Méthodes Modernes de la Mécanique Céleste (Goutelas ’89)*, edited by D. Benest and C. Froeschlé (Editions Frontières, Gif-sur-Yvette, France, 1989), pp. 285–329.
- [40] Y. G. Sinai, “Construction of Markov partitions”, *Funct. Anal. Appl.* **2**, 245–253 (1968).

- [41] S. Smale, “Differentiable dynamical systems”, *Bull. Amer. Math. Soc.* **73**, 747–817 (1967).

Figure 15.16: The stable/unstable manifolds of the equilibrium $(x_q, x_q) = (0, 0)$ of 2-dimensional flow (15.12).



15.7 Examples

Example 15.1. A simple stable/unstable manifolds pair. Consider the 2-dimensional ODE system

$$\frac{dx}{dt} = -x, \quad \frac{dy}{dt} = y + x^2, \quad (15.12)$$

The flow through a point $x(0) = x_0, y(0) = y_0$ can be integrated

$$x(t) = x_0 e^{-t}, \quad y(t) = (y_0 + x_0^2/3) e^t - x_0^2 e^{-2t}/3. \quad (15.13)$$

Linear stability of the flow is described by the stability matrix

$$\mathbf{A} = \begin{pmatrix} -1 & 0 \\ 2x & 1 \end{pmatrix}. \quad (15.14)$$

The flow is hyperbolic, with a real expanding/contracting eigenvalue pair $\lambda_1 = 1, \lambda_2 = -1$, and area preserving. The right eigenvectors at the point (x, y)

$$\mathbf{e}^{(1)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \mathbf{e}^{(2)} = \begin{pmatrix} 1 \\ -x \end{pmatrix}. \quad (15.15)$$

can be obtained by acting with the projection operators (see *example A4.2 Decomposition of 2-dimensional vector spaces*)

$$P_i = \frac{\mathbf{A} - \lambda_j \mathbf{1}}{\lambda_i - \lambda_j} : \quad P_1 = \begin{bmatrix} 0 & 0 \\ x & 1 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 1 & 0 \\ -x & 0 \end{bmatrix} \quad (15.16)$$

on an arbitrary vector. Matrices P_i are orthonormal and complete.

The flow has a degenerate pair of equilibria at $(x_q, y_q) = (0, 0)$, with eigenvalues (stability exponents), $\lambda_1 = 1, \lambda_2 = -1$, eigenvectors $\mathbf{e}^{(1)} = (0, 1), \mathbf{e}^{(2)} = (1, 0)$. The unstable manifold is the y axis, and the stable manifold is given by (see figure 15.16)

$$y_0 + \frac{1}{3}x_0^2 = 0 \Rightarrow y(t) + \frac{1}{3}x(t)^2 = 0. \quad (15.17)$$

(N. Lebovitz)

[return to p. 292](#)

Example 15.2. A section at a fixed point with a complex Floquet multiplier pair.

The simplest choice of a Poincaré section for a fixed (or periodic) point x_q with a complex Floquet multiplier pair is the plane $\mathcal{P}$ specified by the fixed point (located at the tip of the vector x_q) and the eigenvector $\text{Im } \mathbf{e}^{(k)}$ perpendicular to the plane. A point x is in the section $\mathcal{P}$ if it satisfies the condition

$$(x - x_q) \cdot \text{Im } \mathbf{e}^{(k)} = 0. \quad (15.18)$$

In the neighborhood of x_q , the spiral out/in motion is in the $\{\text{Re } \mathbf{e}^{(k)}, \text{Im } \mathbf{e}^{(k)}\}$ plane and, thus, is guaranteed to be cut by the Poincaré section $\mathcal{P}$ normal to $\mathbf{e}^{(k)}$.

[return to p. 293](#)

Example 15.3. Recoding 3-disk dynamics in binary. (Continued from example 14.1) The $\mathcal{A} = \{1, 2, 3\}$ symbolic dynamics for the 3-disk system is neither unique, nor necessarily the smartest one. Before proceeding, it pays to quotient the symmetries of the dynamics in order to obtain a more efficient description. We will do this in a quick way here and redo it later in more detail in sect. 15.5.

As the three disks are equidistantly spaced, the disk labels are arbitrary; what is important is how a trajectory evolves as it hits subsequent disks, not what label the starting disk had. We exploit this symmetry by *recoding*, in this case replacing the absolute disk labels by relative symbols, indicating the type of collision. For the 3-disk game of pinball, there are two topologically distinct kinds of collisions, figure 1.8 (e):

$$s_i = \begin{cases} 0 & : \text{pinball returns to the disk it came from} \\ 1 & : \text{pinball continues to the third disk.} \end{cases} \quad (15.19)$$

exercise 14.1

exercise 15.7

In the *binary* recoding of 3-disk symbolic dynamics, the prohibition of self-bounces is automatic. If the disks are sufficiently far apart, there are no further restrictions on symbols, the symbolic dynamics is complete, and *all* binary sequences (see table 18.1) are admissible.

exercise 14.2

It is intuitively clear that as we go backward in time (reverse the velocity vector), we also need increasingly precise specification of $x_0 = (s_0, p_0)$ in order to follow a given past itinerary. Another way to look at the survivors after two bounces is to plot $\mathcal{M}_{s_1, s_2}$, the intersection of $\mathcal{M}_{s_2}$ with the strips $\mathcal{M}_{s_1}$, obtained by time reversal (the velocity changes sign $\sin \phi \rightarrow -\sin \phi$). $\mathcal{M}_{s_1, s_2}$ is a ‘rectangle’ of nearby trajectories (see figure 15.3 (a)), which have arrived from disk s_1 and are heading for disk s_2 . (continued in example 15.5)

return to p. 296

Example 15.4. A Hénon repeller complete horseshoe. (Continued from example 3.5) Consider 2-dimensional *Hénon map*

exercise 3.5

$$(x_{n+1}, y_{n+1}) = (1 - ax_n^2 + by_n, x_n). \quad (15.20)$$

If you start with a small ball of initial points centered around the fixed point x_0 and iterate the map, the ball will be stretched and squashed along the unstable manifold W_0^u . Iterated backward in time,

$$(x_{n-1}, y_{n-1}) = (y_n, -b^{-1}(1 - ay_n^2 - x_n)), \quad (15.21)$$

this small ball of initial points traces out the stable manifold W_0^s . Their intersections enclose the region $\mathcal{M}$, figure 15.4 (a). Any point outside the W_0^s border of $\mathcal{M}$ escapes to infinity forward in time, while –by time reversal– any point outside W_0^u border arrives from infinity back in time. In this way the unstable/stable manifolds define the topologically, invariant and optimal initial region $\mathcal{M}$; all orbits that stay confined for all times are confined to $\mathcal{M}$.

The Hénon map models qualitatively the Poincaré section return map of figure 14.7 (b). For $b = 0$ the Hénon map reduces to a parabola (14.20), as shown in sects. 3.3 and 34.1. For $b \neq 0$ it is a slightly fattened parabola; by construction, it takes a rectangular initial area and returns it bent as a horseshoe. Parameter a controls the amount of stretching, while parameter b controls the amount of compression of the folded horseshoe. For clarity, fix the parameter values to $a = 6$, $b = -1$; the map is then strongly stretching but area preserving, the furthest away from the strongly dissipative examples discussed in sect. 14.2. The map is quadratic, so it has 2 fixed points $x_0 = f(x_0)$, $x_1 = f(x_1)$ indicated in figure 15.4 (a). For the parameter values at hand, they are both unstable.

Iterated one step forward, the region $\mathcal{M}$ is stretched and folded into a *Smale* horseshoe drawn in figure 15.4(b). Label the two forward intersections $f(\mathcal{M}) \cap \mathcal{M}$ by $\mathcal{M}_s$, with $s \in \{0, 1\}$. The horseshoe consists of the two strips $\mathcal{M}_0, \mathcal{M}_1$ and also the bent segment that lies entirely outside the W_0^s line. Since all points in this segment escape to infinity under forward iteration, this region can safely be cut out and thrown away.

Iterated one step backwards, the region $\mathcal{M}$ is again stretched and folded into a horseshoe, figure 15.4(c). As stability and instability are interchanged under time reversal, this horseshoe is transverse to the forward one. Once again, the points in the horseshoe bend and wander off to infinity as $n \rightarrow -\infty$, and we are then left with the two (past) strips $\mathcal{M}_0, \mathcal{M}_1$. Iterating two steps forward we obtain the four strips $\mathcal{M}_{11}, \mathcal{M}_{01}, \mathcal{M}_{00}, \mathcal{M}_{10}$, and iterating backwards we obtain the four strips $\mathcal{M}_{00}, \mathcal{M}_{01}, \mathcal{M}_{11}, \mathcal{M}_{10}$ transverse to the forward ones just as for the 3-disk pinball game figure 1.8(e). Iterating three steps forward we get 8 strips, and so on *ad infinitum*. For a detailed discussion of the Hamiltonian, area-preserving Hénon map repeller symbolic dynamics sketched above, see Li and Tomsovic [34]. (continued in example 15.9)

[return to p. 296](#)

Example 15.5. Recoding ternary symbolic dynamics in binary. Given a ternary sequence and labels of 2 preceding disks, rule (15.19) fixes the subsequent binary symbols. Here we list an arbitrary ternary itinerary, and the corresponding binary sequence:

$$\begin{array}{ll} \text{ternary} & : \quad 3 \ 1 \ 2 \ 1 \ 3 \ 1 \ 2 \ 3 \ 2 \ 1 \ 2 \ 3 \ 1 \ 3 \ 2 \ 3 \\ \text{binary} & : \quad \cdot \ 1 \ 0 \ 1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 0 \end{array} \quad (15.22)$$

The first 2 disks initialize the trajectory and its direction; $3 \mapsto 1 \mapsto 2 \mapsto \dots$. Due to the 3-disk symmetry the six distinct 3-disk sequences, initialized by 12, 13, 21, 23, 31, 32 respectively, have the same weights and the same size state space partitions. They are also both coded by a single binary sequence. (continued in example 15.7)

[return to p. 303](#)

Example 15.6. C_2 recoded. Assume that each orbit is uniquely labeled by an infinite string $\{s_i\}$, $s_i \in \{+, -\}$ and that the dynamics is C_2 -equivariant under the $+$ $\leftrightarrow$ $-$ interchange. Periodic orbits separate into two classes, the self-dual configurations $+-$, $++--$, $+++---$, $+-+-+-$, $+-+-+-$, $\dots$, with multiplicity $m_p = 1$, and the pairs $+-$, $-$, $++-$, $--+$, $\dots$, with multiplicity $m_p = 2$. For example, as there is no absolute distinction between the ‘left’ or the ‘right’ lobe of the Lorenz attractor, figure 3.5(a), the Floquet multipliers satisfy $\Lambda_+ = \Lambda_-$, $\Lambda_{++-} = \Lambda_{+--}$, and so on.

[exercise 11.3](#)

[exercise 25.6](#)

The symmetry reduced labeling $\rho_i \in \{0, 1\}$ is related to the full state space labeling $s_i \in \{+, -\}$ by

$$\begin{array}{ll} \text{If } s_i = s_{i-1} & \text{then } \rho_i = 1 \\ \text{If } s_i \neq s_{i-1} & \text{then } \rho_i = 0 \end{array} \quad (15.23)$$

For example, the fixed point $\overline{+} = \dots + + + \dots$ maps into $\dots 111 \dots = \overline{1}$, and so does the fixed point $\overline{-}$. The 2-cycle $\overline{+-} = \dots - + - + \dots$ maps into fixed point $\dots 000 \dots = \overline{0}$, and the 4-cycle $\overline{++--} = \dots - - + + - - + + \dots$ maps into 2-cycle $\dots 0101 \dots = \overline{01}$. A list of such reductions is given in table 15.1.

[return to p. 303](#)

Example 15.7. D_3 recoded - 3-disk game of pinball. (Continued from example 15.5) The D_3 recoding can be worked out by a glance at figure 15.12 (a) (Continuation of example 11.11). For the symmetric 3-disk game of pinball, the fundamental domain is bounded by a disk segment and the two adjacent sections of the symmetry axes that act as mirrors (see figure 15.12 (b)). The three symmetry axes divide the space into six copies of the fundamental domain. Any trajectory on the full space can be pieced together from bounces in the fundamental domain, with symmetry axes replaced by flat mirror reflections. The binary $\{0, 1\}$ reduction of the ternary three disk $\{1, 2, 3\}$ labels has a simple geometric interpretation, figure 1.8 (e). A collision of type 0 reflects the projectile to the disk it comes from (back-scatter); whereas, after a collision of type 1, the projectile continues to the third disk. For example, $\overline{23} = \dots 232323 \dots$ maps into $\dots 000 \dots = \overline{0}$ (and so do $\overline{12}$ and $\overline{13}$), $\overline{123} = \dots 12312 \dots$ maps into $\dots 111 \dots = \overline{1}$ (and so does $\overline{132}$), and so forth. Such reductions for short cycles are given in table 15.2, figure 15.12 and figure 11.6.

return to p. 304

Example 15.8. Pinball game, Poincaré dissected. (Continued from sect. 1.4 and chapter 9) A phase-space orbit is fully specified by its position and momentum at a given instant, so no two distinct phase-space trajectories can intersect. The configuration space trajectories, however, can and do intersect, in rather unilluminating ways, as e.g. in figure 15.14 (d). Also, it can be rather hard to perceive the systematics of orbits from their configuration space shapes. The problem is that we are looking at the projections of 4-dimensional state space trajectories onto a 2-dimensional configuration subspace. A much clearer picture of the dynamics is obtained by constructing a set of Poincaré sections.

Suppose that the pinball has just bounced off disk 1. Depending on its position and outgoing angle, it could proceed to either disk 2 or 3. Not much happens in between the bounces—the ball just travels at constant velocity along a straight line—so we can reduce the 4-dimensional flow to a 2-dimensional map $P_{\sigma_k \leftarrow \sigma_j}$ that maps the coordinates (Poincaré section $\mathcal{P}_1$) of the pinball from one disk edge to another. Just after the moment of impact the trajectory is defined by s_n , the arc-length position of the n th bounce along the billiard wall, and $p_n = p \sin \phi_n$ the outgoing momentum component parallel to the billiard wall at the point of impact, figure 15.15 (a). These coordinates (due to Birkhoff) are smart, as they conserve the phase-space volume. Trajectories originating from one disk can hit either of the other two disks, or escape. We label the survivor state space regions $\mathcal{P}_{12}$, $\mathcal{P}_{13}$. In terms of the three Poincaré sections, one for each disk, the dynamics is reduced to the set of six maps

$$(s_{n+1}, p_{n+1}) = P_{\sigma_{n+1} \leftarrow \sigma_n}(s_n, p_n), \quad \sigma \in \{1, 2, 3\} \quad (15.24)$$

from the boundary of a disk to the boundary of the next disk, figure 15.15 (b). The explicit form of this map is easily written down, see example 9.1, but much more economical is the symmetry quotiented version of chapter 11 which replaces the above 6 forward maps by a *return map* pair P_0, P_1 .

exercise 9.7

return to p. 307

Example 15.9. A Hénon repeller subshift. (Continued from example 15.4) The Hénon map acts on the binary partition as a shift map. Figure 15.5 illustrates action $f(\mathcal{M}_0) = \mathcal{M}_0$. The square $[01.01]$ gets mapped into the rectangles $\sigma[01.01] = [10.1] = \{[10.10], [10.11]\}$, see figure 15.4 (e). Further examples can be gleaned from figure 15.4.

return to p. 299

Exercises

15.1. **A Smale horseshoe.** The Hénon map of example 3.5

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 - ax^2 + by \\ x \end{bmatrix} \quad (15.25)$$

maps the $[x, y]$ plane into itself - it was constructed by Hénon [28] in order to mimic the Poincaré section of once-folding map induced by a flow like the one sketched in figure 14.7. For definitiveness, fix the parameters to $a = 6$, $b = -1$.

- Draw a rectangle in the (x, y) plane such that its n th iterate by the Hénon map intersects the rectangle 2^n times.
 - Construct the inverse of the matrix in (15.25).
 - Iterate the rectangle back in the time. How many intersections are there between the n forward and m backward iterates of the rectangle?
 - Use the above information about the intersections to guess the (x, y) coordinates for the two fixed points, a 2-periodic point, and points on the two distinct 3-cycles from table 18.1. The exact periodic points are computed in exercise 16.11.
- 15.2. **A simple stable/unstable manifolds pair.** Integrate flow (15.12), verify (15.13). Check that the projection matrices P_i (15.16) are orthonormal and complete. Use them to construct right and left eigenvectors; check that they are mutually orthogonal. Explain why (15.17) is the equation for the stable manifold. (N. Lebovitz)
- 15.3. **Kneading Danish pastry.** Write down the $(x, y) \rightarrow (x, y)$ mapping that implements the baker's map

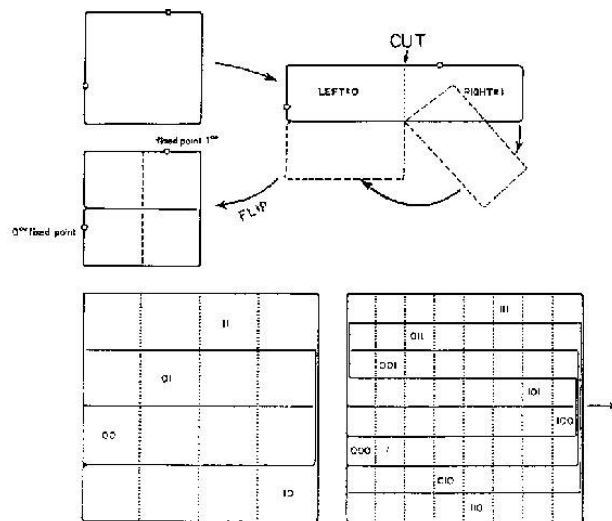


FIG. 4. Iterative construction of the symbol plane.

Figure: *Kneading danish pastry: symbol square representation of an orientation reversing once-folding map obtained by fattening the Smale horseshoe intersections of figure 15.4 into a unit square. In the symbol square, the dynamics maps rectangles into rectangles by a decimal point shift. together with the inverse mapping. Sketch a few rectangles in symbol square and their forward and backward images. (Hint: the mapping is very much like the tent map (14.21)).*

- 15.4. **Kneading danish without flipping.** The baker's map of exercise 15.3 includes a flip - a map of this type is called an orientation reversing once-folding map. Write down the $(x, y) \rightarrow (x, y)$ mapping that implements an orientation preserving baker's map (no flip; Jacobian determinant = 1). Sketch and label the first few folds of the symbol square.
- 15.5. **Orientation reversing once-folding map.** By adding a reflection around the vertical axis to the horseshoe map g we get the orientation reversing map $\tilde{g}$ shown in the second Figure above. $\tilde{Q}_0$ and $\tilde{Q}_1$ are oriented as Q_0 and Q_1 , so the definition of the future topological coordinate γ is identical to the γ for the orientation preserving horseshoe. The inverse intersections $\tilde{Q}_0^{-1}$ and $\tilde{Q}_1^{-1}$ are oriented so that $\tilde{Q}_0^{-1}$ is opposite to Q , while $\tilde{Q}_1^{-1}$ has the same orientation as Q . Check that the past topological coordinate δ is given by

$$w_{n-1} = \begin{cases} 1 - w_n & \text{if } s_n = 0 \\ w_n & \text{if } s_n = 1 \end{cases}, \quad w_0 = s_0$$

$$\delta(x) = 0.w_0w_{-1}w_{-2}\dots = \sum_{n=1}^{\infty} w_{1-n}/2^n. \quad (15.26)$$

- 15.6. **Infinite symbolic dynamics.** Let σ be a function that returns zero or one for every infinite binary string, $\sigma : \{0, 1\}^{\mathbb{N}} \rightarrow \{0, 1\}$. Its value is represented by $\sigma(\epsilon_1, \epsilon_2, \dots)$ where the ϵ_i are either 0 or 1. We will now define an operator $\mathcal{T}$ that acts on observables on the space of binary strings. A function a is an observable if it has bounded variation, that is, if

$$\|a\| = \sup_{\{\epsilon_i\}} |a(\epsilon_1, \epsilon_2, \dots)| < \infty.$$

For these functions

$$\begin{aligned} \mathcal{T}a(\epsilon_1, \epsilon_2, \dots) &= a(0, \epsilon_1, \epsilon_2, \dots)\sigma(0, \epsilon_1, \epsilon_2, \dots) \\ &\quad + a(1, \epsilon_1, \epsilon_2, \dots)\sigma(1, \epsilon_1, \epsilon_2, \dots). \end{aligned}$$

- (a) (easy) Consider a finite version T_n of the operator $\mathcal{T}$:

$$\begin{aligned} T_n a(\epsilon_1, \epsilon_2, \dots, \epsilon_{1,n}) &= \\ &= a(0, \epsilon_1, \epsilon_2, \dots, \epsilon_{n-1})\sigma(0, \epsilon_1, \epsilon_2, \dots, \epsilon_{n-1}) + \\ &+ a(1, \epsilon_1, \epsilon_2, \dots, \epsilon_{n-1})\sigma(1, \epsilon_1, \epsilon_2, \dots, \epsilon_{n-1}). \end{aligned}$$

Show that T_n is a $2^n \times 2^n$ matrix. Show that its trace is bounded by a number independent of n .

- (b) (medium) With the operator norm induced by the function norm, show that $\mathcal{T}$ is a bounded operator.
- (c) (hard) Show that $\mathcal{T}$ is not trace class.

- 15.7. **3-disk fundamental domain cycles.** (continued from exercise 11.3) Try to sketch $\bar{0}$, $\bar{1}$, $\bar{01}$, $\bar{001}$, $\bar{011}$, $\dots$ in the fundamental domain, and interpret the symbols $\{0, 1\}$ by relating them to topologically distinct types of collisions. Compare with table 15.2. Then try to sketch the location of periodic points in the Poincaré section of the billiard flow. The point of this exercise is that, while in the configuration space, longer cycles look like a hopeless jumble. However, in the Poincaré section they are clearly and logically ordered. The Poincaré section

is always to be preferred to projections of a flow onto the configuration space coordinates, or any other subset of state space coordinates which does not respect the topological organization of the flow.

- 15.8. **3-disk pruning.** (Not easy) Show that for 3-disk game of pinball the pruning of orbits starts at $R : a = 2.04821419 \dots$, figure 14.6. (K.T. Hansen)

Chapter 16

Fixed points, and how to get them

Cycles. Is there anything they can't do?
— Mason Porter, channeling Homer Simpson

HAVING SET UP the dynamical context, we now turn to the key and unavoidable numerical task in this subject; we must search for the solutions (x, T) , $x \in \mathbb{R}^d$, $T \in \mathbb{R}_+$ satisfying the *periodic orbit condition*

$$\begin{aligned} f^T(x) &= x, & T > 0, & \quad (\text{flow}) \\ f^n(x) &= x, & n \geq 1, & \quad (\text{map}) \end{aligned} \tag{16.1}$$

for a given flow or map.

In chapters 21 and 22 we will establish that spectra of evolution operators can be extracted from periodic orbit sums:

$$\sum (\text{spectral eigenvalues}) = \sum (\text{periodic orbits}) .$$

Hence, periodic orbits are the necessary ingredient for evaluation of the spectra of evolution operators. We need to know what periodic orbits can exist, and the symbolic dynamics developed so far is an invaluable tool toward this end.

This chapter, a continuation of chapter 7, is intended as a hands-on guide to extracting periodic orbits, and a reader should skip this chapter on a first reading. You can return to it whenever the need for finding actual cycles arises. A serious cyclist will want to also learn about the variational methods to find cycles, chapter 34. They are particularly useful when little is known about the topology of a flow, such as in high-dimensional periodic orbit searches.

chapter 34



fast track:
chapter 17, p. 335

Due to the exponential divergence of nearby trajectories in chaotic dynamical systems, fixed point searches based on direct solutions of the fixed-point condition (16.1) as an initial value problem can be numerically very unstable. Methods that start with initial guesses for a number of points along the cycle are considerably more robust and safer. Examples of this approach are the multipoint shooting method, described in sect. 16.2, and the variational methods of chapter 34. If you want to do billiards, proceed straight to sect. 34.3.

chapter 34

A prerequisite for any exhaustive cycle search is a good understanding of the topology of the flow. A preliminary step to any serious periodic orbit calculation is preparing a list of all distinct admissible prime periodic symbol sequences, such as the list given in table 18.1. Relations between temporal symbol sequences and the spatial layout of topologically distinct regions in the state space (see chapters 14 and 15) should enable us to guess the location of a series of periodic points along a cycle. Armed with such an informed guess, we proceed to improve it by methods such as the Newton-Raphson iteration. We show how this works by applying the Newton method to 1- and d -dimensional maps. But first, where are the cycles?

16.1 Where are the cycles?

Q: What if you choose a really bad initial condition and it doesn't converge? A: Well then you only have yourself to blame.

— T.D. Lee

The simplest and conceptually easiest setting for guessing where the cycles are is the case of planar billiards. The Maupertuis principle of least action here dictates that the physical trajectories minimize the length of an approximate orbit that visits a desired sequence of boundary bounces.



example 16.1
p. 330

If we were only so lucky. Real life finds us staring at something like Yang-Mills or Navier-Stokes equations, utterly clueless. What to do?

One, there is always mindless computation. Some might be satisfied with any rampaging robot that finds “the most important” cycles. The ergodic explorations of recurrences sometimes perform admirably well, and we discuss this next.

16.1.1 Cycles from long time series

Two wrongs don't make a right, but three lefts do.

—Appliance guru

(L. Rondoni and P. Cvitanović)

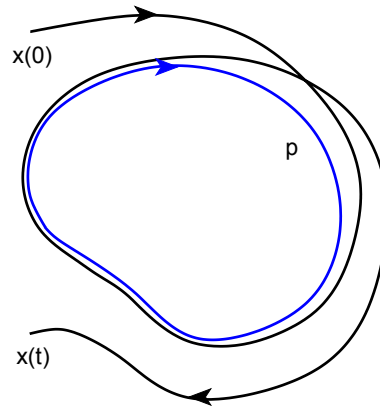


Figure 16.1: An ergodic trajectory can shadow an unstable periodic orbit p for a finite time.

The equilibria and periodic orbits (with the exception of sinks and stable limit cycles) are never seen in simulations and experiments because they are unstable. Nevertheless, one does observe close passes to the least unstable equilibria and periodic orbits, as in figure 16.1. Ergodic exploration by long-time trajectories (or long-lived transients, in case of strange repellers) can uncover state space regions with near finite time recurrences. In addition, such trajectories preferentially sample the natural measure of the ‘turbulent’ flow, and searches initiated within the state space concentrations of natural measure will bias the search toward dynamically important invariant solutions.

remark 16.1

section 19.1

The search consists of following a long trajectory in state space, and looking for close returns of the trajectory to itself, see figure 16.1. Whenever the trajectory almost closes in a loop (within a given tolerance), a point close to this near miss of a cycle can be taken as an initial condition. Supplemented by a Newton routine described below, a sequence of improved initial conditions may indeed rapidly lead to closing a cycle. The method preferentially finds the least unstable orbits, while missing the more unstable ones that contribute little to the cycle expansions.

This blind search is seriously flawed. In contrast to the 3-disk example 16.1, it is not systematic. It gives no insight into the organization of the ergodic sets, and this blind search can easily miss very important cycles. Foundations to a systematic exploration of ergodic state space are laid in chapters 14 and 15, but they are a bit of work to implement.

16.1.2 Cycles found by thinking

Thinking is extra price.
—Dicho Colombiano

A systematic charting out of state space starts out by a hunt for equilibrium points. If the equations of motion are a finite set of ODEs, setting the velocity field $v(x)$ in (2.9) to zero reduces the search for equilibria to a search for zeros of a set of algebraic equations. We should be able, in principle, to enumerate and determine all real and complex zeros in such cases, e.g. the Lorenz example 2.2 and the Rössler example 2.3. If the equations of motion and boundary conditions are

invariant under some symmetry, some equilibria can be determined by symmetry considerations: if a function is e.g. antisymmetric, it must vanish at the origin, e.g. the Lorenz $EQ_0 = (0, 0, 0)$ equilibrium.

In regard to other equilibria (if you have no better idea), create a state space grid, about 50 grid points across $\mathcal{M}$ in each dimension, and then compute a velocity field $v_k = v(x_k)$ at each grid point x_k . By using this approach, a few million v_k values are easily stored. Plot x_k for which $|v_k|^2 < \epsilon$, $\epsilon \ll |v_{max}|^2$ but sufficiently large that a few thousand x_k are plotted. If the velocity field varies smoothly across the state space, the regions $|v_k|^2 < \epsilon$ isolate the (candidate) equilibria. Start a Newton iteration with the smallest $|v_k|^2$ point within each region. Barring exceptionally fast variations in $v(x)$, this should yield all equilibrium points.

For ODEs equilibria are fixed points of algebraic sets of equations, but steady states of PDEs such as the Navier-Stokes flow are themselves solutions of ODEs or PDEs, and much harder to determine.

Equilibria—by definition—do not move, so they cannot be “turbulent.” What makes them dynamically important are their stable/unstable manifolds. A chaotic trajectory can be thought of as a sequence of visitations to equilibrium neighborhoods. Typically such neighborhoods have many stable, contracting directions and a handful of unstable directions. Our strategy will be to generalize the billiard Poincaré section maps $P_{s_{n+1} \leftarrow s_n}$ of example 15.8 to maps from a section of the unstable manifold of equilibrium s_n to the section of stable manifold of equilibrium s_{n+1} . Thus, this reduces the continuous time flow to a sequence of maps. These Poincaré section maps do double duty, providing us both with an exact representation of dynamics in terms of maps and with a review of symbolic dynamics.

We showed in the Lorenz flow example 14.4 how to reduce the 3-dimensional Lorenz flow to a 1-dimensional return map. In the Rössler flow example 2.3, we sketched the attractor by running a long chaotic trajectory and noting that the attractor is very thin. Otherwise the return maps that we plotted were disquieting, since figure 3.4 does not appear to be a 1-to-1 map. In the next example, we show how to use such information to locate cycles approximately. Additionally, in the remainder of this chapter and in chapter 34, we shall learn how to turn such guesses into highly accurate cycles.

16.2 Multipoint shooting method

(F. Christiansen)

Periodic orbits of period n are fixed points of f^n , so in principle we could use the simple Newton method described above to find them. However, this is not an optimal strategy. The function f^n oscillates wildly, with as many as 2^n or more closely spaced fixed points, and finding a specific periodic point, such as one with a given symbolic sequence, requires a *very* good starting guess. For binary symbolic dynamics we must expect to improve the accuracy of our initial guesses

by at least a factor of 2^n to find orbits of period n . Furthermore, the Jacobian of f^n can be ill-conditioned because its matrix elements can grow like Λ^n , where Λ is the leading multiplier of a single discrete time step Jacobian. A better alternative is the *multiplier* or *multiple shooting method*, with the Jacobian matrix broken up into a product of single-step Jacobian matrices, each with eigenvalues $\approx \Lambda$. While it might be very hard to give a precise initial guess for a long periodic orbit, a rough guess for each point along the desired trajectory might suffice, if our guesses are informed by a good state space partition. As for individual short trajectory segments, the errors have no time to explode exponentially. That is why in chapter 14 we have developed a qualitative theory of how these cycle points are laid out topologically.

For a 1-dimensional map, a cycle of period n is a zero of the n -dimensional vector function F :

$$F(x) = F \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 - f_n \\ x_2 - f_1 \\ \vdots \\ x_n - f_{n-1} \end{pmatrix}, \quad f_n = f(x_n). \quad (16.2)$$

The relationship between temporal symbol sequences and the spatial layout of topologically distinct regions of state space (see chapter 14) enables us to guess the location of a series of periodic points along a cycle. Armed with such informed initial guesses, we can initiate a Newton-Raphson iteration. The iteration in Newton's method now takes the form

$$\frac{d}{dx} F(x)(x' - x) = -F(x), \quad f'_n = f'(x_n)$$

where $\frac{d}{dx} F(x)$ is an $[n \times n]$ matrix:

$$\frac{d}{dx} F(x) = \begin{pmatrix} 1 & & & & -f'_n \\ -f'_1 & 1 & & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & & \\ & & & 1 & \\ & & & -f'_{n-1} & 1 \end{pmatrix}. \quad (16.3)$$

This matrix can easily be explicitly inverted as shown in example 16.2. You would think we were done, but not really. The explicit answer involves exponentially large factors, such as the n -cycle stability multiplier $J = f'_n \cdots f'_3 f'_2 f'_1$ and can easily lead to overflows for longer cycles.

 example 16.2
p. 330

When one sets up Newton iteration on a computer, it is not necessary to write the left hand side as a matrix. All one needs is a vector containing the $f'(x_i)$'s and a vector containing the n 'th column, i.e., the cumulative product of the $f'(x_i)$'s and a vector containing the right hand side. After iteration the vector containing the right hand side is the correction to the initial guess.

16.2.1 d -dimensional maps



By using clever initial guesses from a system's symbolic dynamics, we can easily extend the Newton-Raphson iteration method to d -dimensional maps. In this case $f'(x_i)$ is a $[d \times d]$ matrix, and $\frac{d}{dx}F(x)$ is an $[nd \times nd]$ matrix. In each of the steps above, we are manipulating d rows of the left-hand-side matrix. (Remember that matrices do not commute - always multiply from the left.) In inverting the n th element of the diagonal, we are inverting a $[d \times d]$ matrix $(1 - \prod f'(x_i))$ which can be done as long as none of the eigenvalues of $\prod f'(x_i)$ equals 1, i.e., if the cycle has no marginally stable eigen-directions.



example 16.3
p. 331

16.3 Cost function

(R. Paškauskas and P. Cvitanović)

It pays to think in terms of a *cost* (or *error*) function $I(\Delta x) = (x + \Delta x - f(x + \Delta x))^2/2$. Periodic orbit condition (16.1) corresponds both to a zero of $I(\Delta x)$, and to its first Δx variation. If we expand $I(\Delta x)$ to the second order in Δx , then $\tilde{I} \approx \tilde{\Delta x}^2/2 + (x - f(x)) \cdot \tilde{\Delta x} + (x - f(x))^2/2$, where $\tilde{\Delta x} = (1 - J(x))\Delta x$. To find an extremum, we set the derivative with respect to $\tilde{\Delta x}$ to zero. As the term $(x - f(x))^2/2$ is a constant under Δx variation, let us define an unconstrained *cost function*

$$I_0(\tilde{\Delta x}) = \frac{1}{2} \tilde{\Delta x} \cdot \tilde{\Delta x} + (x - f(x)) \cdot \tilde{\Delta x}, \quad (16.4)$$

Setting the derivative of this function

$$\frac{\partial I_0(\tilde{\Delta x})}{\partial \tilde{\Delta x}} = \tilde{\Delta x} + x - f(x) = (1 - J(x)) \cdot \Delta x + x - f(x) \quad (16.5)$$

to zero recovers the Newton setup (7.3).

Next, we need to enforce the constraint that curbs the directions in which Δx can point. Lagrange multipliers come to help.

A *local surface of section* can be constructed when $f(x)$ is “near” the initial point x . A natural choice is a hyperplane perpendicular to the velocity vector $v(x)$. The reference point x_0 in (7.10) is x itself, and the surface of section condition is $U(x + \Delta x) = v(x) \cdot \Delta x = 0$. By introducing a Lagrange multiplier λ and assembling a cost function with the constraint, we get

$$I_1(\tilde{\Delta x}, \lambda) = \frac{1}{2} \tilde{\Delta x} \cdot \tilde{\Delta x} + [x - f(x)] \cdot \tilde{\Delta x} + \lambda v(x) \cdot \tilde{\Delta x}. \quad (16.6)$$

Now we differentiate $I_1(\Delta x, \lambda)$ with respect to each argument and set the derivatives to zero. We recover the Newton setup (7.7), with the Lagrange multiplier

$\lambda = \Delta t$ interpreted as the time increment needed to place $f(x)$ onto the section, $f(x) \rightarrow f(x) + v(f(x))\Delta t$.

A *global surface of section* is a fixed surface $U(x + \Delta x) - U(x_0) \approx \partial U(x)\Delta x + U(x) - U(x_0)$ that hopefully transects all or a significant portion of recurrent parts of the flow. It is not as ‘natural’ as the local section (7.6) but hard to avoid in practice, if one is interested not only in the fixed point itself but also in the global reach of its unstable manifold. The simplest choice is a hyperplane (7.10). The cost function and the variational equations are then

$$\begin{aligned} I_2(\Delta x, \Delta t) &= \frac{1}{2}\Delta x[1 - J(x)]\Delta x + (x - f(x))\Delta x \\ &+ \Delta t(\partial U(x)\Delta x + U(x) - U(x_0)), \end{aligned} \quad (16.7)$$

$$\begin{bmatrix} 1 - J(x) & \partial U(x) \\ \partial U(x) & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta t \end{pmatrix} = - \begin{pmatrix} x - f(x) \\ U(x) - U(x_0) \end{pmatrix}. \quad (16.8)$$

Further *continuous symmetries* can be handled in the same fashion. Suppose, for example, that we are searching for periodic orbits of a Hamiltonian flow. In this situation, periodic orbits not only have the time-translation symmetry but energy-translation symmetry as well. What is energy-translation symmetry? If there exists a periodic orbit at x with energy $H(x) = E$, and period T , it is very likely that it belongs to a family of orbits $(x + \epsilon\Delta x(E), T + \epsilon\Delta t(E))$ continuous under variation of E . As with time-translation symmetry, this implies a unit Floquet multiplier. Indeed, we know from sect. 8.5 that symplectic eigenvalues come in pairs, so a unit multiplier in the time direction implies a unit multiplier in its dual, the energy direction, $(\Lambda_t, \Lambda_E, \dots) = (1, 1, \dots)$. Extending the number of constraints is no longer a problem, since we can add more Lagrange multipliers. Consider the following system

$$\begin{aligned} I_3(\Delta x, \lambda_1, \lambda_2) &= \Delta x[1 - J(x)]\Delta x/2 + (x - f(x))\Delta x \\ &+ \lambda_1 (U(x + \Delta x) - U(x_0)) + \lambda_2 (H(x + \Delta x) - E_0) \end{aligned} \quad (16.9)$$

$$\begin{bmatrix} 1 - J(x) & \partial U(x) & \partial H(x) \\ \partial U(x) & 0 & 0 \\ \partial H(x) & 0 & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = - \begin{pmatrix} x - f(x) \\ U(x) - U(x_0) \\ H(x) - E_0 \end{pmatrix}. \quad (16.10)$$

This Newton iteration setup searches for a periodic orbit of a Hamiltonian flow with a global surface of section $U(x) = U(x_0)$ and fixed energy E_0 . Note that these instructions do not put every iteration on a surface $U(x) = U(x_0)$ and energy $H(x) = E_0$, unless the surface is a plane $U(x) = a \cdot (x - x_0)$. Instead, this setup assures that the iterations (provided they converge) will approach the surface super-exponentially.

For periodic orbits multi-point shooting generalizes in the same way as (16.3), but with n additional equations – one for each point on a Poincaré section. The Newton setup looks like this:

$$\left[\begin{array}{cccc|cccc} 1 & & & & -J_n & & & \\ -J_1 & 1 & & & & -v_1 & & \\ & \dots & 1 & & & & \ddots & \\ & & \dots & 1 & & & & -v_n \\ & & & -J_{n-1} & 1 & \dots & & \\ \hline & a & & & & 0 & & \\ & & \ddots & & & & \ddots & \\ & & & a & & & & 0 \end{array} \right] \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_n \\ \Delta t_1 \\ \vdots \\ \Delta t_n \end{pmatrix} = \begin{pmatrix} -F_1 \\ -F_2 \\ \vdots \\ -F_n \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (16.11)$$

Solving this equation resembles the corresponding task for maps. However, we will need to invert a $[(d+1)n \times (d+1)n]$ matrix rather than a $[d \times d]$ matrix.

Résumé

A prerequisite for a systematic and complete cycle search is a good (but hard to come by) understanding of the topology of the flow. Usually one starts by - possibly analytic - determination of the equilibria of the flow. Their locations, stabilities, stability eigenvectors and invariant manifolds offer skeletal information about the topology of the flow. The next step is numerical long-time evolution of “typical” trajectories of the dynamical system under investigation. Such numerical experiments build up the “natural measure” and reveal which regions are most frequently visited. Periodic orbit searches can then be initialized by taking nearly recurring orbit segments and deforming them into closed orbits. With a sufficiently good initial guess, Newton-Raphson iteration then yields the period T and the location of a periodic point x_p .

section 19.4.1

The problem one faces with high-dimensional flows is that their topology is hard to visualize, and that even with a decent starting guess for a point on a periodic orbit, methods like the Newton-Raphson method are likely to fail. Methods that start with initial guesses for a number of points along the cycle, such as the multipoint shooting method of sect. 16.2, are more robust. Relaxation (or variational) methods take this strategy to its logical extreme. These methods start with a guess of the entire orbit, not just a guess of a few points along a periodic orbit. Just as these methods are intimately related to variational principles and path integrals, we postpone their introduction until chapter 34.

chapter 34

Commentary

Remark 16.1. Close recurrence searches. For low-dimensional flows (for high-dimensional flows, forget about it), picking initial guesses for periodic orbits from close

recurrences of a long ergodic trajectory seems like an obvious idea. Nevertheless, if you need to cite the paper where the method was first deployed, cite ref. [1]. Such methods have been deployed by many, among them G. Tanner, L. Rondoni, G. Morris, C.P. Dettmann, and R.L. Davidchack [2–5, 7] (see also sect. 23.7). Sometimes one can determine most of the admissible itineraries and their weights without working too hard, but this method comes with no guarantee. For a clear and simple overview of how to compute periodic orbits in higher dimensions (for example, for Navier-Stokes), see Willis [8] lectures on *Equilibria, periodic orbits and computing them*, [arXiv:1908.06730](https://arxiv.org/abs/1908.06730), a very nice, student friendly introduction.

Remark 16.2. Cycles, searches, and symmetries. A few comments about the role of symmetries in actual extraction of cycles. In the N -disk billiard example, a fundamental domain is a sliver of the N -disk configuration space delineated by a pair of adjoining symmetry axes. The flow may further be reduced to a return map on a Poincaré section. While in principle any Poincaré section will do, a natural choice in the present context are crossings of symmetry axes, see example 8.7. In actual numerical integrations, only the last crossing of a symmetry line needs to be determined. The cycle is run in global coordinates, and the group elements associated with the crossings of symmetry lines are recorded; integration is terminated when the orbit closes in the fundamental domain. Periodic orbits with non-trivial symmetry subgroups are particularly easy to find since their points lie on crossings of symmetry lines, see example 8.7.

References

- [1] D. Auerbach, P. Cvitanović, J.-P. Eckmann, G. Gunaratne, and I. Procaccia, “Exploring chaotic motion through periodic orbits”, *Phys. Rev. Lett.* **58**, 2387–2389 (1987).
- [2] J. J. Crofts and R. L. Davidchack, “Efficient detection of periodic orbits in chaotic systems by stabilizing transformations”, *SIAM J. Sci. Comp.* **28**, 1275–1288 (2006).
- [3] J. J. Crofts and R. L. Davidchack, “On the use of stabilizing transformations for detecting unstable periodic orbits in high-dimensional flows”, *Chaos* **19**, 033138 (2009).
- [4] C. P. Dettmann and P. Cvitanović, “Cycle expansions for intermittent diffusion”, *Phys. Rev. E* **56**, 6687–6692 (1997).
- [5] C. P. Dettmann and G. P. Morriss, “Stability ordering; strong field Lorentz gas”, *Phys. Rev. Lett.* **78**, 4201–4204 (1997).
- [6] C. Dong, “Organization of the periodic orbits in the Rössler flow”, *Int. J. Mod. Phys. B* **22**, 1850227 (2018).
- [7] G. P. Morriss and L. Rondoni, “Periodic orbit expansions for the Lorentz gas”, *J. Stat. Phys.* **75**, 553–584 (1994).
- [8] A. P. Willis, *Equilibria, periodic orbits and computing them*, 2019.

16.4 Examples

Example 16.1. Periodic orbits of billiards.

Consider how this works for a 3-disk pinball game of sect. 15.5. Label the three disks with 1, 2 and 3, and associate to every trajectory an *itinerary*, a sequence of labels indicating the order in which the disks are visited, as in figure 15.14. Given the itinerary, you can construct a guess trajectory by taking a point on the boundary of each disk in the sequence, and connecting them by straight lines. Imagine that this is a rubber band wrapped through 3 rings, and shake the band until it shrinks into the physical trajectory, the rubber band of shortest length.

section 15.5

section 1.4

Extremization of a cycle length requires variation of n bounce positions s_i . The computational problem is to find the extremum values of cycle length $L(s)$ where $s = (s_1, \dots, s_n)$, a task that we postpone to sect. 34.3. As an example, the short periods and stabilities of 3-disk cycles computed this way are listed in table 34.1. Other examples are plotted in figure 15.14. It's a no brainer, and millions of such cycles have been computed.

exercise 34.2

exercise 16.11

return to p. 322

Example 16.2. Newton inversion for a 3-cycle.

Let us illustrate how the Newton multiple shooting method (16.3) works step by step for a 3-cycle. The initial setup for a Newton step is:

$$\begin{pmatrix} 1 & 0 & -f'_3 \\ -f'_1 & 1 & 0 \\ 0 & -f'_2 & 1 \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \end{pmatrix} = - \begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix},$$

where $\Delta x_i = x'_i - x_i$ is the correction to our initial guess x_i , and $F_i = x_i - f_{i-1}$ is the error at i th periodic point. Eliminate the sub-diagonal elements by adding f'_1 times the first row to the second row, and then adding f'_2 times the second row to the third row:

$$\begin{pmatrix} 1 & 0 & -f'_3 \\ 0 & 1 & -f'_1 f'_3 \\ 0 & 0 & 1 - J \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \end{pmatrix} = - \begin{pmatrix} F_1 \\ F_2 + f'_1 F_1 \\ F_3 + f'_2 F_2 + f'_2 f'_1 F_1 \end{pmatrix}.$$

The next step is to invert the last element on the diagonal, i.e., divide the third row by $1 - J$, where $J = f'_3 f'_2 f'_1$ is the stability of the cycle (when the Newton iteration has converged). If this element is zero at the periodic orbit, this step cannot work. In this situation, the Newton method is not a good method to find marginally stable cycles. We now have

$$\begin{pmatrix} 1 & 0 & -f'_3 \\ 0 & 1 & -f'_1 f'_3 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \end{pmatrix} = - \begin{pmatrix} F_1 \\ F_2 + f'_1 F_1 \\ \frac{F_3 + f'_2 F_2 + f'_2 f'_1 F_1}{1 - J} \end{pmatrix},$$

Finally we add f'_3 times the third row to the first row and $f'_1 f'_3$ times the third row to the second row:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \end{pmatrix} = - \frac{1}{1 - J} \begin{pmatrix} F_1 + f'_3 F_3 + f'_3 f'_2 F_2 \\ F_2 + f'_1 F_1 + f'_1 f'_3 F_3 \\ F_3 + f'_2 F_2 + f'_2 f'_1 F_1 \end{pmatrix}.$$

The left hand side matrix is now the unit matrix, and the right hand side is an explicit formula for the corrections to our initial guess, cyclic in form, as it should be.

$$\begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \end{pmatrix} = - \frac{1}{1 - J} \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & f'_3 \\ f'_1 & 0 & 0 \\ 0 & f'_2 & 0 \end{pmatrix} + \begin{pmatrix} 0 & f'_3 f'_2 & 0 \\ 0 & 0 & f'_1 f'_3 \\ f'_2 f'_1 & 0 & 0 \end{pmatrix} \right\} \begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix}.$$

With this, we have gone through one Newton iteration. For an unstable cycle, the error gets contracted by the factor $1/(1 - J)$, with the earlier errors amplified by the orbit instability. For example, Δx_3 receives a contribution from two time steps in the past of form $f'_2 f'_1 F_1$.

[return to p. 325](#)

Example 16.3. Newton method for time delay maps. Some d -dimensional maps (such as the Hénon map (3.18)) can be written as 1-dimensional time delay maps of the form

$$f(x_i) = f(x_{i-1}, x_{i-2}, \dots, x_{i-d}). \quad (16.12)$$

In this case, $\frac{d}{dx} F(x)$ is an $[n \times n]$ matrix as in the case of the usual 1-dimensional maps but with non-zero matrix elements on d off-diagonals.

[return to p. 326](#)

Exercises

16.1. Ulam map periodic points. (continued from exercise 14.8)

- (a) compute the five periodic points of cycle $\overline{10011}$ for the Ulam map (14.22) $f(x) = 4x(1 - x)$, using your Newton or other routine.
- (b) compute the five periodic points of cycle $\overline{10000}$
- (c) plot the above two cycles on the graph of the Ulam map, verify that their topological ordering is as in the ‘canonical’ full tent map exercise 14.8.
- (d) (optional) This works only for the Ulam map: compute periodic points by conjugating the full tent map periodic points of exercise 14.8 using exercise A2.3.

16.2. Cycles stabilities for the Ulam map (exact). In exercise 16.1 you should have observed that the numerical results for the cycle Floquet multipliers (4.43) are exceptionally simple. The Floquet multiplier of the $x_0 = 0$ fixed point is 4, while the eigenvalue of any other n -cycle is $\pm 2^n$. Prove this. (Hint: the Ulam map can be conjugated to the tent map (14.21). This problem is perhaps too hard, but give it a try - the answer is in many introductory books on nonlinear dynamics.)

16.3. Newton-Raphson method. Implement the Newton-Raphson method in 2 dimensions, and apply it to the determination of pinball cycles.

16.4. Cycle stability. Add to the pinball simulator of exercise 9.1 a routine that evaluates the expanding eigenvalue for a given cycle.

16.5. Pinball cycles. Determine the stability and period of all fundamental domain prime cycles of the binary symbol string lengths up to 5 (or longer) for $R : a = 6$ 3-disk pinball.

16.6. Fundamental domain fixed points. Use the formula (9.10) for billiard Jacobian matrix to compute the periods T_p and the expanding eigenvalues Λ_p of the fundamental domain $\bar{0}$ (the 2-cycle of the complete 3-disk space) and $\bar{1}$ (the 3-cycle of the complete 3-disk space) fixed points:

$$\begin{array}{c|cc}
 & T_p & \Lambda_p \\
 \hline
 \bar{0}: & R - 2 & R - 1 + R\sqrt{1 - 2/R} \\
 \bar{1}: & R - \sqrt{3} & -\frac{2R}{\sqrt{3}} + 1 - 2R\sqrt{3}\sqrt{1 - \sqrt{3}/R}
 \end{array} \quad (16.13)$$

We have set the disk radius to $a = 1$.

16.7. Fundamental domain 2-cycle. Verify that for the $\bar{10}$ -cycle the cycle period and the trace of the Jacobian matrix are given by

$$\begin{aligned}
 L_{10} &= 2\sqrt{R^2 - \sqrt{3}R + 1 - 2}, \\
 \text{tr } \mathbf{J}_{10} &= \Lambda_{10} + 1/\Lambda_{10} \\
 &= 2L_{10} + 2 + \frac{1}{2} \frac{L_{10}(L_{10} + 2)^2}{\sqrt{3}R/2 - 1}.
 \end{aligned} \quad (16.14)$$

The $\bar{10}$ -cycle is drawn in figure 15.12. The unstable eigenvalue Λ_{10} follows from (8.36).

- 16.8. **A test of your pinball simulator: $\overline{10}$ -cycle.** Test your exercise 9.4 pinball simulator stability evaluation by checking numerically the exact analytic $\overline{10}$ -cycle stability formula (16.14).
- 16.9. **Rössler flow cycles.** (continuation of exercise 7.1) Determine all cycles for the Rössler flow (2.30), as well as their stabilities, up to 5 Poincaré section returns (Hint: implement (16.3), the multipoint shooting methods for flows; you can cross-check your shortest cycles against the ones listed in the table.) You might find a comparison to Dong [6] *Organization of the periodic orbits in the Rössler flow* helpful.

Table: The Rössler flow (2.30): The itinerary p , a periodic point $x_p = (0, y_p, z_p)$ and the expanding eigenvalue Λ_p for all cycles up to topological period 7.
(J. Mathiesen, G. Simon, A. Basu)

n_p	p	y_p	z_p	Λ_e
1	1	6.091768	1.299732	-2.403953
2	01	3.915804	3.692833	-3.512007
3	001	2.278281	7.416481	-2.341923
	011	2.932877	5.670806	5.344908
4	0111	3.466759	4.506218	-16.69674
5	01011	4.162799	3.303903	-23.19958
	01111	3.278914	4.890452	36.88633
6	001011	2.122094	7.886173	-6.857665
	010111	4.059211	3.462266	61.64909
	011111	3.361494	4.718206	-92.08255
7	0101011	3.842769	3.815494	77.76110
	0110111	3.025957	5.451444	-95.18388
	0101111	4.102256	3.395644	-142.2380
	0111111	3.327986	4.787463	218.0284

- 16.10. **Colinear helium cycles.** Determine the stability and period of all fundamental domain prime cycles up to symbol sequence length 5 or longer for collinear helium of figure 8.2.
- 16.11. **Uniqueness of unstable cycles.**  Prove that there exists only one 3-disk prime cycle for a given finite admissible prime cycle symbol string. Hints: look at the return maps. Can you show that there is exponential contraction to a unique periodic point with a given itinerary? Exercise 34.1 might be helpful in this effort.
- 16.12. **Newton setups for flows.**
- We have formulated three Newton setups for flows: the ‘local’ setup (7.7), the ‘hyperplane’ setup (7.11), and the ‘global’ setup (16.8). Derive (16.8) and verify that if the surface of section is a hyperplane, it reduces to (7.11). (Hint: it is not inconceivable that (7.11) is wrong as it stands.)
 - (optional) Derive (16.10), the Newton setup for Hamiltonian flows.

Part II

Chaos rules

Q UNADRY: all these cycles, but what to do with them? What you have now is a topologically invariant road map of the state space, with the chaotic region pinned down by a rigid skeleton, a tree of *cycles* (periodic orbits) of increasing periods and self-similar structure. In chapter 18 we shall turn this topological dynamics into a multiplicative operation on the state space partitions by means of transition matrices of chapter 17, the simplest examples of evolution operators. This will enable us to *count* the distinct orbits, and in the process touch upon all the main themes of this book, going the whole distance from diagnosing chaotic dynamics to computing zeta functions.

1. Partition the state space and describe all allowed ways of getting from ‘here’ to ‘there’ by means of transition graphs (transition matrices). These generate the totality of admissible itineraries (chapter 17)
2. Learn to count (chapter 18)
3. Learn how to measure what’s important (chapter 19)
4. Learn how to evolve the measure, compute averages (chapter 20)
5. Learn what a ‘Fourier transform’ is for a nonlinear world (chapter 21),
6. and how the short-time / long-time duality is encoded by spectral determinant expression for its spectrum in terms of periodic orbits (chapter 22)
7. Learn how to use short period cycles to describe chaotic world at times much beyond the Lyapunov time (chapter 23)
8. What is all this hard work good for? Deterministic diffusion and foundations of ‘far from equilibrium’ statistical mechanics, for example (chapter 24)
9. Back to hard work: ponder how symmetries simplify spectral determinants (chapter 25 and chapter 26)

Chapter 17

Walkabout: Transition graphs

I think I'll go on a walkabout
find out what it's all about [...] take a ride to the other side
—Red Hot Chili Peppers, 'Walkabout'

IN CHAPTERS 14 AND 15 we learned that invariant manifolds partition the state space in invariant ways and how to name distinct orbits. We have established and related the *temporally* and *spatially* ordered topological dynamics for a class of 'stretch & fold' dynamical systems, and we've also discussed pruning of inadmissible trajectories.

Here we shall use these results to generate the totality of admissible itineraries. This task will be particularly easy for repellers with complete Smale horseshoes and for subshifts of finite type. The admissible itineraries are generated by finite transition matrices, and the topological dynamics can be visualized by means of finite transition graphs. We shall then turn topological dynamics into a linear multiplicative operation on the state space partitions by means of transition matrices, the simplest examples of 'evolution operators.' They will enable us – in chapter 18 – to *count* the distinct orbits.



17.1 Matrix representations of topological dynamics



The allowed transitions between the regions of a partition $\{\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_m\}$ are encoded in the $[m \times m]$ -dimensional transition matrix whose elements take values

$$T_{ij} = \begin{cases} 1 & \text{if the transition } \mathcal{M}_j \rightarrow \mathcal{M}_i \text{ is possible} \\ 0 & \text{otherwise.} \end{cases} \quad (17.1)$$

The transition matrix is an explicit linear representation of topological dynamics. If the partition is a dynamically invariant partition constructed from sta-



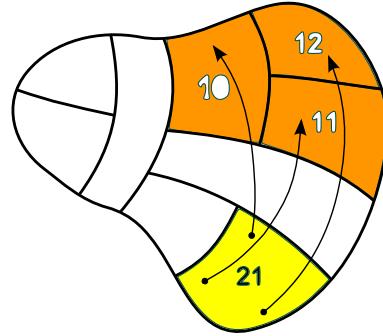


Figure 17.1: Points from region $\mathcal{M}_{21}$ reach regions $\{\mathcal{M}_{10}, \mathcal{M}_{11}, \mathcal{M}_{12}\}$ in one time step. Labeling exemplifies the ‘shift map’ of example 14.5 and (14.12).

ble/unstable manifolds, it encodes the topological dynamics as an invariant law of motion. The allowed transitions at any instant are independent of the trajectory history and, thus, require no memory of the past.

Several related matrices as well will be needed in what follows. Often it is convenient to distinguish between two or more paths connecting the same two regions. This can be encoded by the *adjacency* matrix with non-negative integer entries,

$$A_{ij} = \begin{cases} k & \text{if a transition } \mathcal{M}_j \rightarrow \mathcal{M}_i \text{ is possible in } k \text{ ways} \\ 0 & \text{otherwise.} \end{cases} \quad (17.2)$$

More generally, we shall encounter $[m \times m]$ matrices which assign different real or complex weights to different transitions,

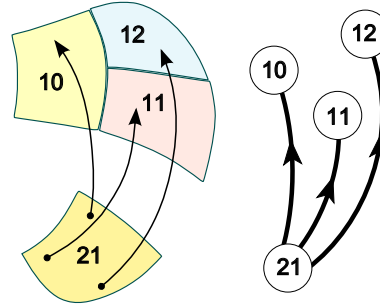
$$L_{ij} = \begin{cases} L_{ij} \in \mathbb{R} \text{ or } \mathbb{C} & \text{if } \mathcal{M}_j \rightarrow \mathcal{M}_i \text{ is allowed} \\ 0 & \text{otherwise.} \end{cases} \quad (17.3)$$

As in statistical physics, we shall refer to these as *transfer* matrices.

$\mathcal{M}_i$ is *accessible* from $\mathcal{M}_j$ in k steps if $(L^k)_{ij} \neq 0$. A matrix L is called *reducible* if there exists one or more index pairs $\{i, j\}$ such that $(L^k)_{ij} = 0$ for all k , otherwise the matrix is *irreducible*. This means that a trajectory starting in any partition region eventually reaches all of the partition regions, i.e., the partition is dynamically transitive or indecomposable, as assumed in (2.5). The notion of topological *transitivity* is crucial in ergodic theory. A mapping is transitive if it has a dense orbit. If that is not the case, state space decomposes into disconnected pieces, each of which can be analyzed separately by a separate irreducible matrix. Region $\mathcal{M}_i$ is said to be *transient* if no trajectory returns to it. Region $\mathcal{M}_j$ is said to be *absorbing* if no trajectory leaves it, $L_{jj} \neq 0$, $L_{ij} = 0$ for all $i \neq j$. Hence it suffices to restrict our considerations to irreducible matrices.

If L has strictly positive entries, $L_{ij} > 0$, the matrix is called *positive*; if $L_{ij} \geq 0$, the matrix is called *non-negative*. Matrix L is said to be *eventually positive* or *Perron-Frobenius*, if L^k is positive for some power k (as a consequence, the matrix is transitive as well). A non-negative matrix whose columns conserve probability, $\sum_i L_{ij} = 1$, is called a *Markov, probability* or *stochastic* matrix.

Figure 17.2: Topological dynamics: shrink each state space partition region in figure 17.1 to a *node*, and indicate the possibility of reaching a region by a *directed link*. The links stand for transition matrix elements $T_{10,21} = T_{11,21} = T_{12,21} = 1$; remaining $T_{ij,21} = 0$.



A subshift (14.15) of finite type is a *topological dynamical system* (Σ, σ) , where the shift σ acts on the space of all admissible itineraries (s_k)

$$\Sigma = \{(s_k)_{k \in \mathbb{Z}} : T_{s_{k+1}s_k} = 1 \text{ for all } k\}, \quad s_k \in \{a, b, c, \dots, z\}. \quad (17.4)$$

The task of generating the totality of admissible itineraries is particularly easy for subshifts of finite type. These admissible itineraries are generated by finite transition matrices, and the topological dynamics can be visualized by means of finite transition graphs.

17.2 Transition graphs: wander from node to node

Let us extract from a state space partition, such as figure 17.1, its topological essence. First, indicate a partition region M_a by a *node*, and then indicate the possibility of reaching the region M_b , $L_{ba} \neq 0$ by a *directed link*, as in figure 17.2. Do this for all nodes. The result is a *transition graph*.

A *transition graph* consists of a set of *nodes*, one for each letter in the alphabet $\mathcal{A} = \{a, b, c, \dots, z\}$, connected by a set of directed *edges*. A directed link starts out from node j and terminates at node i , whenever the matrix element (17.3) takes value $L_{ij} \neq 0$. A link connects two nodes, or it originates and terminates on the same node (a ‘self-loop’). For example, if a partition includes regions labeled $\{\dots, M_{101}, M_{110}, \dots\}$, the transition matrix element connecting the two is drawn as $L_{101,110} = \textcircled{101} \leftarrow \textcircled{110}$, whereas $L_{0,0} = \textcircled{\cdot} \textcircled{\ominus}$. Here a dotted link indicates that the shift $\sigma(x_{011\dots}) = x_{11\dots}$ involves symbol 0, and a full one a shift $\sigma(x_{110\dots}) = x_{10\dots}$ that involves 1. A $j \rightarrow \dots \rightarrow k$ walk (*path*, *itinerary*) traverses a connected set of directed links, starting at node j and ending at node k . A *loop* (*periodic orbit*, *cycle*) is a walk that ends at the starting node (which can be any node along the loop), for example

$$t_{011} = L_{110,011} L_{011,101} L_{101,110} = \textcircled{101} \leftarrow \textcircled{110} \leftarrow \textcircled{011} \leftarrow \textcircled{101} \quad (17.5)$$

Our convention for ordering indices is that the successive steps in a visitation sequence $j \rightarrow i \rightarrow k$ are generated by matrix multiplication from the left, $T_{kj} = \sum T_{ki} T_{ij}$. Two graphs are *isomorphic*, if one can be obtained from the other by relabeling links and nodes. As we are interested in recurrent (transitive, indecomposable) dynamics, we restrict our attention to *irreducible* or *strongly connected*

graphs, i.e., graphs for which there is a path from any node to any other node. (In a *connected* graph one may reach node j from node k , but not node k from node j .)

A transition graph compactly describes the ways in which the state space regions map into each other, accounts for finite memory effects in dynamics, and generates the totality of admissible trajectories as the set of all possible walks along its links. Construction of a good transition graph is, like combinatorics, unexplainable (check page 268). The only way to learn is by some diagrammatic gymnastics, so we recommend that you work your way through the examples and exercises in lieu of a plethora of baffling definitions.



example 17.1
p. 344



example 17.2
p. 344



example 17.3
p. 344

The complete unrestricted symbolic dynamics is too simple to be illuminating, so we turn next to the simplest example of pruned symbolic dynamics, which is the finite subshift obtained by prohibition of repeats of one of the symbols; let us say $\_11\_$. This situation arises, for example, in billiard dynamics and also circle maps, where this kind of symbolic dynamics describes ‘golden mean’ rotations (we shall return to this example in chapter 32).



example 17.4
p. 344



example 17.5
p. 345

exercise 18.6
exercise 18.8

In the complete N -ary symbolic dynamics case (see example 17.2) the choice of the next symbol requires no memory of the previous ones. However, any further refinement of the state space partition requires finite memory.



example 17.6
p. 345

For M -step memory the only nonvanishing matrix elements are of the form $T_{s_1 s_2 \dots s_{M+1}, s_0 s_1 \dots s_M}$, $s_{M+1} \in \{0, 1\}$. This is a sparse matrix, as the only non vanishing entries in the $a = s_0 s_1 \dots s_M$ column of T_{ba} are in the rows $b = s_1 \dots s_{M+1}$ and $b = s_1 \dots s_M 1$. If we increase the number of remembered steps, the transition matrix grows large quickly, as the N -ary dynamics with M -step memory requires an $[N^{M+1} \times N^{M+1}]$ matrix. Since the matrix is very sparse, it pays to find a compact representation for T . Such a representation is afforded by transition graphs, which are not only compact, but also give us an intuitive picture of the topological dynamics.

exercise 18.1

17.3 Transition graphs: stroll from link to link

(P. Cvitanović and Matjaž Gomilšek)

What do finite graphs have to do with infinitely long trajectories? To understand the main idea, let us construct an infinite rooted tree graph that explicitly enumerates all possible itineraries. In this construction the nodes are unlabeled, and the

links labeled (or colored, or dotted in different ways), signifying different kinds of transitions.

A *tree graph* is an undirected graph (its links have no sense of direction, $j \leftrightarrow i$) in which there exists exactly one path between any two of its nodes. A tree graph is thus connected (irreducible) and contains no loops; i.e., it is not possible to return to any of its nodes by a walk along a sequence of distinct links. A *rooted tree graph* is a directed graph (its links are directed, $j \rightarrow i$), obtained from an undirected tree graph by picking a distinguished node, called the *root*, and orienting all links in the tree so that they point away from the root.

Each node in a directed graph has an *in-degree* (number of links pointing towards it, or the number of ‘parents’), and an *out-degree* (number of links pointing away from it, or the number of ‘children’). An *internal node* has both in- and out-degree ≥ 1 . In a rooted tree graph, all nodes have exactly one parent (in-degree = 1), except for the root, which is the single “parentless” node (in-degree = 0), with all links pointing away from it. An *external node (leaf)* is a “childless” node, with in-degree ≥ 1 , out-degree = 0. We shall refer to a node with known ancestors, but as yet unspecified descendants, as a *free node*.



example 17.7
p. 345

We illustrate how trees are related to transition graphs by first working out the simplest example of pruned symbolic dynamics, the finite subshift obtained by prohibition of repeats of one of the symbols, let us say `_00_`. As we shall see, for finite grammars a rooted tree (and, by extension, but less obviously, the associated transition graph) is the precise statement of what is meant topologically by a “self-similar” fractal; supplemented by scaling information, such a rooted tree generates a self-similar fractal. Any slightly more complicated grammar merits a full section of its own, here sect. 17.3.1.



example 17.8
p. 346

17.3.1 Converting pruning blocks into transition graphs



Suppose now that, by hook or crook, you have been so lucky fishing for pruning rules that you now know the grammar (14.16) in terms of a finite set of pruning blocks $\mathcal{G} = \{b_1, b_2, \dots, b_k\}$, of lengths $\leq m$. Our task is to generate all admissible itineraries. What to do?

We have already seen the main ingredients of a general algorithm: (1) the transition graph encodes the self-similarities of the tree of all itineraries, and (2) if we have a pruning block of length m , we need to descend m levels before we can start identifying the self-similar sub-trees.

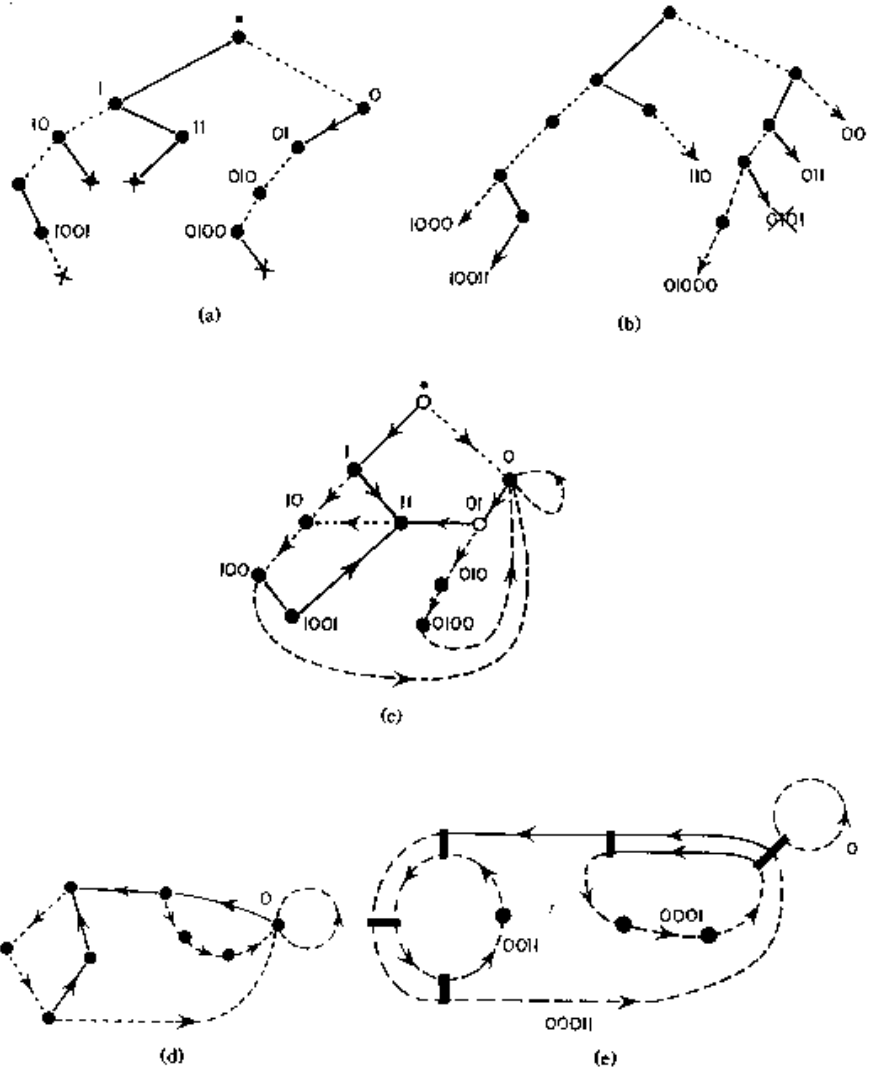


Figure 17.3: Conversion of the pruning front of figure 15.11 (b) into a finite transition graph. (a) Starting with the root node “.”, delineate all pruning blocks on the binary tree. A solid line stands for “1” and a dashed line for “0”. The ends of forbidden strings (i.e., the external nodes) are marked with $\times$. Label all internal nodes by reading the bits connecting “.”, the root of the tree, to the node. (b) Indicate all admissible starting blocks by arrows. (c) Recursively drop the leading bits in the admissible blocks; if the truncated string corresponds to an internal node in (a), identify them. (d) Delete the transient, non-circulating nodes; all admissible sequences are generated as walks on this finite transition graph. (e) Identify all distinct non-intersecting (products of) loops and construct the determinant (18.33).

Finite grammar transition graph algorithm.

1. Starting with the root of the tree, delineate all branches that correspond to all pruning blocks; implement the pruning by removing the last node in each pruning block (marked ‘ $\times$ ’ in figure 17.3 (a)).
2. Label all nodes internal to pruning blocks by the itinerary connecting the root node to the internal node, figure 17.3 (a). Why? So far we have pruned forbidden branches by looking m_b steps into the future for a given pruning block, let’s say $b = 10110$. However, the blocks with the right combination of past and future $[1.0110]$, $[10.110]$, $[101.10]$ and $[1011.0]$ are also pruned. In other words, any node whose near past coincides with the beginning of a pruning block is potentially dangerous - a branch further down the tree might get pruned.
3. Add to each remaining node, including the root, all remaining branches allowed by the alphabet, and label their top (free) nodes, figure 17.3 (b). Why? Each one of the free nodes is the beginning point of an infinite tree,

a tree that should be similar to another one originating closer to the root of the whole tree.

4. Check that the labels of the newly added free nodes do not themselves contain any pruning blocks. If they do, remove them (marked ‘×’ in figure 17.3 (b)).
5. Pick one of the remaining free nodes (e.g. closest to the root of the entire tree), forget the most distant symbol in its past. Does the truncated itinerary correspond to an internal node? If yes, identify the two nodes. If not, forget the next symbol in its past, repeat. If no such truncated past corresponds to any internal node, identify with the root of the tree.
This is a little bit abstract, so let’s say the free node in question is [1010.]. Three time steps into the past is [010.]. That is not dangerous, as no pruning block in this example starts with 0. Now forget the third step in the past: [10.] is dangerous, as that is the start of the pruning block [10.110]. Hence the free node [1010.] should be identified with the internal node [10.].
6. Repeat until all free nodes have been tied back into internal nodes or the root.
7. Clean up: check whether every node can be reached from every other node. Remove the transient nodes, i.e., the nodes to which dynamics never returns.
8. The result is a transition graph. There is no guarantee that this is the smartest, most compact transition graph possible for a given pruning (if you have a better algorithm, teach us), but walks around it do generate all admissible itineraries, and nothing else.



example 17.9
p. 346

Résumé

The set of all admissible itineraries is generated multiplicatively by transition matrices, diagrammatically by transition graphs. Pruning rules for inadmissible sequences are implemented by constructing corresponding transition matrices and/or transition graphs. These matrices are the simplest examples of evolution operators, prerequisite to developing a theory of averaging over chaotic flows. The initial chapters 2 to 4 focus on things local, such as a representative point, a short-time trajectory, and a neighborhood. In this chapter and the next one, we make a courageous leap and go global.

Commentary

Remark 17.1. Transition graphs. We enjoyed studying Lind and Marcus [15] and Adler [1] introductions to symbolic dynamics, Markov partitions and transition graphs.

Alligood, Sauer and Yorke [2] discussion of baker's maps, Smale horseshoes and their symbolic dynamics is simple and clear. Consult also Kitchens [14] and Rodriguez-Gonzalez [16]. Finite transition graphs or finite automata are discussed in refs. [5, 11, 17]. They belong to the category of regular languages. Transition graphs for unimodal maps are discussed in refs. [8, 10, 13]. For a deep dive into the physics history of uses of transition graphs, consult ref. [9]. (see also remark 14.1)

Remark 17.2. Inflating transition graphs. In the above examples the symbolic dynamics has been encoded by labeling links in the transition graph. Alternatively one can encode the dynamics by labeling the nodes, as in example 17.6, where the 4 nodes refer to 4 Markov partition regions $\{M_{00}, M_{01}, M_{10}, M_{11}\}$ and the 8 links refer to the 8 non-zero entries in the 2-step memory transition matrix (17.11).

Remark 17.3. The unbearable growth of transition graphs. A construction of finite Markov partitions is described in refs. [3, 4, 12], as well as in innumerable many other references.

If two regions in a Markov partition are not disjoint but share a boundary, the boundary trajectories require special treatment in order to avoid overcounting, see sect. 25.4.3. If the image of a trial partition region cuts across only a part of another trial region and thus violates the Markov partition condition (14.2), further refinement of the partition is needed to distinguish distinct trajectories.

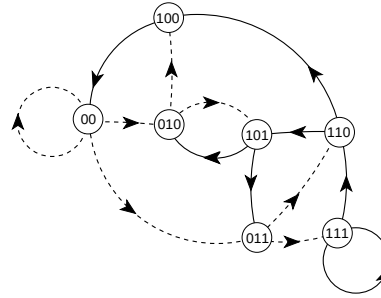
The finite transition graph construction sketched above is not necessarily the minimal one; for example, the transition graph of figure 17.3 does not generate only the “fundamental” cycles (see chapter 23), but shadowed cycles as well, such as t_{00011} in (18.33). For methods of reduction to a minimal graph, consult refs. [7, 8, 10]. Furthermore, when one implements the time reversed dynamics by the same algorithm, one usually gets a graph of a very different topology, even though both graphs generate the same admissible sequences and have the same determinant. The algorithm described here makes some sense for 1-dimensional dynamics, but is unnatural for 2-dimensional maps that are treated as 1-dimensional maps. In practice, generic pruning grows longer and longer, and more plentiful pruning rules. For generic flows the refinements might never stop, and we almost always have to deal with infinite Markov partitions, such as those that will be discussed in sect. 18.5. Not only do the transition graphs get more and more unwieldy, they have the unpleasant property that every time we add a new rule, the graph has to be constructed from scratch. The graph might even look very different from the previous one, even though it leads to a minute modification of the topological entropy. The most determined effort to construct such graphs may be in ref. [6]. Still, this is the best technology available, until the day when a reader alerts us to something superior.

References

- [1] R. L. Adler, “Symbolic dynamics and Markov partitions”, *Bull. Amer. Math. Soc.* **35**, 1–56 (1998).
- [2] K. T. Alligood, T. D. Sauer, and J. A. Yorke, *Chaos, An Introduction to Dynamical Systems* (Springer, New York, 1996).

- [3] A. Boyarsky, “A matrix method for estimating the Liapunov exponent of one-dimensional systems”, *J. Stat. Phys.* **50**, 213–229 (1988).
- [4] A. Boyarsky and M. Skarowsky, “On a class of transformations which have unique absolutely continuous invariant measures”, *Trans. Amer. Math. Soc.* **255**, 243–262 (1979).
- [5] D. M. Cvektović, M. Doob, and H. Sachs, *Spectra of Graphs* (Academic, New York, 1980).
- [6] G. D’Alessandro, P. Grassberger, S. Isola, and A. Politi, “On the topology of the Hénon map”, *J. Phys. A* **23**, 5285–5294 (1990).
- [7] P. Grassberger, “Toward a quantitative theory of self-generated complexity”, *Int. J. Theor. Phys.* **25**, 907–938 (1986).
- [8] P. Grassberger, “On symbolic dynamics of one-humped maps of the interval”, *Z. Naturf. A* **43**, 671–680 (1988).
- [9] P. Grassberger, *Some comments on computational mechanics, complexity measures, and all that*, 2017.
- [10] P. Grassberger, R. Badii, and A. Politi, “Scaling laws for invariant measures on hyperbolic and nonhyperbolic attractors”, *J. Stat. Phys.* **51**, 135–178 (1988).
- [11] J. E. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages and Computation* (Addison-Wesley, Reading MA, 1979).
- [12] C. S. Hsu and M. C. Kim, “Construction of maps with generating partitions for entropy evaluation”, *Phys. Rev. A* **31**, 3253–3265 (1985).
- [13] S. Isola and A. Politi, “Universal encoding for unimodal maps”, *J. Stat. Phys.* **61**, 263–291 (1990).
- [14] B. P. Kitchens, *Symbolic Dynamics: One-sided, Two-sided and Countable State Markov Shifts* (Springer, Berlin, 1997).
- [15] D. A. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding* (Cambridge Univ. Press, Cambridge UK, 1995).
- [16] A. Rodriguez-Gonzalez, *Qualitative Dynamics and Chaos* (2023).
- [17] A. Salomaa, *Formal Languages* (Academic, San Diego, 1973).

Figure 17.4: Transition graph (graph whose links correspond to the nonzero elements of a transition matrix T_{ba}) describes which regions b can be reached from region a in one time step. The 7 nodes correspond to the 7 regions of partition (17.8). The links represent non-vanishing transition matrix elements, such as $T_{101,110} = \textcircled{1}$. Dotted links correspond to a shift by symbol 0, and the solid-line links correspond to a shift by symbol 1.



17.4 Examples

Example 17.1. Full binary shift. Consider a full shift on two-state partition $\mathcal{A} = \{0, 1\}$, with no pruning restrictions. The transition matrix and the corresponding transition graph are

$$T = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{array}{c} \textcircled{0} \quad \textcircled{1} \\ \text{---} \quad \text{---} \\ \text{---} \quad \text{---} \end{array} \quad (17.6)$$

Dotted links correspond to shifts originating in region 0, and solid-line links correspond to shifts originating in region 1. The admissible itineraries are generated as walks on this transition graph. (continued in example 17.7)

[return to p. 338](#)

Example 17.2. Complete N -ary dynamics. If all transition matrix entries equal unity (one can reach any region from any other region in one step),

$$T_c = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}, \quad (17.7)$$

the symbolic dynamics is called *complete*, or a *full shift*. The corresponding transition graph is obvious, but a bit tedious to draw for arbitrary N .

[return to p. 338](#)

Example 17.3. A 7-state transition graph. Consider a state space partitioned into 7 regions

$$\{\mathcal{M}_{00}, \mathcal{M}_{011}, \mathcal{M}_{010}, \mathcal{M}_{110}, \mathcal{M}_{111}, \mathcal{M}_{101}, \mathcal{M}_{100}\}. \quad (17.8)$$

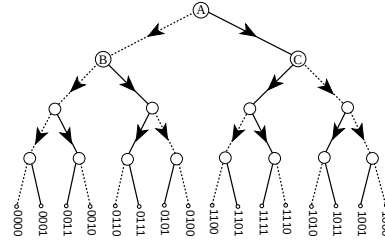
Let the evolution in time map the regions into each other by acting on the labels as shift (15.7): $\mathcal{M}_{011} \rightarrow \{\mathcal{M}_{110}, \mathcal{M}_{111}\}$, $\mathcal{M}_{00} \rightarrow \{\mathcal{M}_{00}, \mathcal{M}_{011}, \mathcal{M}_{010}\} \dots$, with nonvanishing $L_{110,011}, L_{011,00}, \dots$, etc.. This is compactly summarized by the transition graph of figure 17.4. (continued as example 18.6)

[return to p. 338](#)

Example 17.4. Pruning rules for a 3-disk alphabet. As the disks are convex, there can be no two consecutive reflections off the same disk. Hence, the covering symbolic dynamics consists of all sequences which include no symbol repetitions 11, 22, 33. This is a finite set of finite length *pruning rules*, where the dynamics is a subshift of finite type (see (14.16) for definition). In this example, the transition matrix and graph is given by

[exercise 18.1](#)

Figure 17.5: The self-similarity of the complete binary symbolic dynamics represented by a rooted binary tree. Trees originating in nodes $B, C, \dots$ (actually - any node) are the same as the tree originating in root node A . Level $m = 4$ partition is labeled by 16 binary strings, coded by dotted (0) and full (1) links read down the tree, starting from A . See also figure 14.10.



$$T = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{array}{c} \text{graph with 3 nodes and directed edges} \end{array}. \quad (17.9)$$

return to p. 338

Example 17.5. ‘Golden mean’ pruning. Consider a subshift on two-state partition $\mathcal{A} = \{0, 1\}$, with the simplest grammar $\mathcal{G}$ possible, a single pruned block $b = \_11\_$ (consecutive repeat of symbol 1 is inadmissible). The state $\mathcal{M}_0$ maps both onto $\mathcal{M}_0$ and $\mathcal{M}_1$, but the state $\mathcal{M}_1$ maps only onto $\mathcal{M}_0$. The transition matrix and the corresponding transition graph are

$$T = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{array}{c} \text{graph with 2 nodes and directed edges} \end{array}. \quad (17.10)$$

Admissible itineraries correspond to walks on this finite transition graph. (continued in example 17.8)

return to p. 338

Example 17.6. Finite memory transition graphs. For the binary labeled repeller with complete binary symbolic dynamics, we might chose to partition the state space into four regions $\{\mathcal{M}_{00}, \mathcal{M}_{01}, \mathcal{M}_{10}, \mathcal{M}_{11}\}$, a 1-step refinement of the initial partition $\{\mathcal{M}_0, \mathcal{M}_1\}$. Such partitions are drawn in figure 15.3, as well as figure 1.8. Topologically f acts as a left shift (15.7), and its action on the rectangle $[.01]$ is to move the decimal point to the right, to $[0.1]$, forget the past, $[.1]$, and land in either of the two rectangles $\{[.10], [.11]\}$. By filling in the matrix elements for the other three initial states, we obtain the 1-step memory transition matrix/graph acting on the 4-regions partition

exercise 14.7

$$T = \begin{bmatrix} T_{00,00} & 0 & T_{00,10} & 0 \\ T_{01,00} & 0 & T_{01,10} & 0 \\ 0 & T_{10,01} & 0 & T_{10,11} \\ 0 & T_{11,01} & 0 & T_{11,11} \end{bmatrix} = \begin{array}{c} \text{graph with 4 nodes and directed edges} \end{array}. \quad (17.11)$$

(continued in example 18.7)

return to p. 338

Example 17.7. Complete binary topological dynamics. Mark a dot ‘.’ on a piece of paper. That will be the root of our tree. Draw two short directed lines out of the dot, end each with a dot. The full line will signify that the first symbol in an itinerary is ‘1,’ and the dotted line will signify ‘0.’ Repeat the procedure for each of the two new dots, and then for the four dots, and so on. The result is the binary tree of figure 17.5. Starting at the top node, the tree enumerates exhaustively all distinct finite itineraries of lengths $n = 1, 2, 3, \dots$

$\{0, 1\} \quad \{00, 01, 10, 11\}$
 $\{000, 001, 010, 011, 100, 101, 111, 110\} \dots$

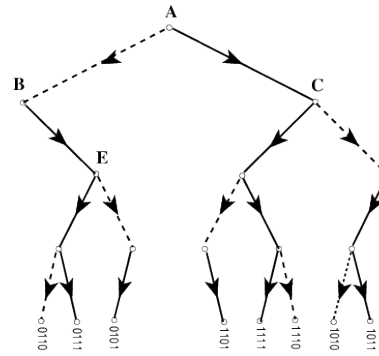


Figure 17.6: The self-similarity of the `_00_` pruned binary tree, trees originating from nodes *C* and *E* are the same as the entire tree.

The $n = 4$ nodes in figure 17.5 correspond to 16 distinct binary strings of length 4, and so on. By habit we have drawn the tree as the alternating binary tree of figure 14.10, but that has no significance as far as enumeration of itineraries is concerned - a binary tree with labels in the natural order, as increasing binary ‘decimals’ would serve just as well.

The trouble with an infinite tree is that it does not fit on a piece of paper. On the other hand, we are not doing much - at each node we are turning either left or right. Hence all nodes are equivalent. In other words, the tree is self-similar; the trees originating in nodes *B* and *C* are themselves copies of the entire tree. The result of identifying $B = A$, $C = A$ is a single node, 2-link transition graph with adjacency matrix (17.2)

$$A = \begin{bmatrix} 2 \end{bmatrix} = \text{graph with one node } A=B=C \text{ and two self-loops.} \quad (17.12)$$

An itinerary generated by the binary tree figure 17.5, no matter how long, corresponds to a walk on this graph. This is the most compact encoding of the complete binary symbolic dynamics. Any number of more complicated transition graphs, such as the 2-node (17.6) and 4-node (17.11) graphs, generate all itineraries as well and might sometimes be preferable.

exercise 18.6
exercise 18.5
return to p. 339

Example 17.8. ‘Golden mean’ pruning. (a link-to-link version of example 17.5) Now the admissible itineraries are enumerated by the pruned binary tree of figure 17.6. Identification of nodes $A = C = E$ leads to the finite 2-node, 3-links transition graph

$$T = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} = \text{graph with two nodes } B \text{ and } A=C=E. \quad (17.13)$$

As 0 is always followed by 1, the walks on this graph generate only the admissible itineraries. This is the same graph as the 2-node graph (17.10), with full and dotted lines interchanged. (continued in example 18.4)

return to p. 339

Example 17.9. Heavy pruning.

We complete this training by examples by implementing the pruning of figure 15.11 (b). The pruning blocks are

$$[100.10], [10.1], [010.01], [011.01], [11.1], [101.10]. \quad (17.14)$$

Blocks 01101, 10110 contain the forbidden block 101, so they are redundant as pruning rules. Draw the *pruning tree* as a section of a binary tree with 0 and 1 branches and label

each internal node by the sequence of 0's and 1's connecting it to the root of the tree, figure 17.3 (a). These nodes are potentially dangerous - beginnings of blocks that might end up pruned. Add the side branches to those nodes, figure 17.3 (b). As we continue down such branches we have to check whether the pruning imposes constraints on the sequences so generated. We do this by knocking off the leading bits and checking whether the shortened strings coincide with any of the internal pruning tree nodes: $00 \rightarrow 0$; $110 \rightarrow 10$; $011 \rightarrow 11$; $0101 \rightarrow 101$ (pruned); $1000 \rightarrow 00 \rightarrow 00 \rightarrow 0$; $10011 \rightarrow 0011 \rightarrow 011 \rightarrow 11$; $01000 \rightarrow 0$.

The trees originating in identified nodes are identical, so the tree is "self-similar." Now connect the side branches to the corresponding nodes, figure 17.3 (d). Nodes "." and 1 are transient nodes; no sequence returns to them, and as you are interested here only in infinitely recurrent sequences, delete them. The result is the finite transition graph of figure 17.3 (d); the admissible bi-infinite symbol sequences are generated as all possible walks on this graph.

[return to p. 341](#)

Exercises

- 17.1. **Time reversibility.**  Hamiltonian flows are time reversible. Does that mean that their transition graphs are symmetric in all node $\rightarrow$ node links, their transition matrices are adjacency matrices, symmetric and diagonalizable, and that they have only real eigenvalues?
- 17.2. **Alphabet {0,1}, prune _1000_, _00100_, _01100_.** This example is motivated by the pruning front description of the symbolic dynamics for the Hénon-type maps, see remark 15.3.
- step 1.** _1000_ prunes all cycles with a _000_ subsequence with the exception of the fixed point $\bar{0}$; hence we factor out $(1 - t_0)$ explicitly, and prune _000_ from the rest. This means that x_0 is an isolated fixed point - no cycle stays in its vicinity for more than 2 iterations. In the notation of sect. 17.3.1, the alphabet is $\{1, 2, 3; \bar{0}\}$, and the remaining pruning rules have to be rewritten in terms of symbols $2=10, 3=100$:
- step 2.** alphabet $\{1, 2, 3; \bar{0}\}$, prune _33_, _213_, _313_. This means that the 3-cycle $\bar{3} = \overline{100}$ is pruned and no long cycles stay close enough to it for a single _100_ repeat. Prohibition of _33_ is implemented by dropping the symbol "3" and extending the alphabet by the allowed blocks 13, 23:
- step 3.** alphabet $\{1, 2, \underline{13}, \underline{23}; \bar{0}\}$, prune _213_, _23 13_, _13 13_, where $\underline{13} = 13, \underline{23} = 23$ are now used as single letters. Pruning of the repetitions $\underline{13} \underline{13}$ (the 4-cycle $\overline{13} = \overline{1100}$ is pruned) yields the
- result:** alphabet $\{1, 2, \underline{23}, \underline{113}; \bar{0}\}$, unrestricted 4-ary dynamics. The other remaining possible blocks $\underline{213}_, \underline{2313}_$ are forbidden by the rules of step 3.

Chapter 18

Counting

I'm gonna close my eyes
And count to ten
I'm gonna close my eyes
And when I open them again
Everything will make sense to me then

—Tina Dico, ‘Count To Ten’

WE ARE NOW in a position to apply the periodic orbit theory to the first and the easiest problem in theory of chaotic systems, which is cycle counting. This is the simplest illustration of the *raison d’être* of periodic orbit theory. Here we derive a duality transformation that relates *local* information to *global* averages. *Local* information relates to the next admissible symbol in a symbol sequence, and *global* averages refer to the mean rate of growth of the number of cycles with increasing cycle period. In chapter 17 we have transformed, by means of the transition matrices / graphs, the topological dynamics of chapter 14 into a multiplicative operation. Here we show that the n th power of a transition matrix counts all itineraries of length n . The asymptotic growth rate of the number of admissible itineraries is therefore given by the leading eigenvalue of the transition matrix; the leading eigenvalue is in turn given by the leading zero of the characteristic determinant of the transition matrix, which is - in this context - called the *topological zeta function*.

For flows with finite transition graphs this determinant is a finite *topological polynomial* which can be read off the graph. However, (a) even something as humble as the quadratic map generically requires an infinite partition (sect. 18.5), but (b) the finite partition approximants converge exponentially fast.

This method goes well beyond the problem at hand and forms the core of the entire treatise, making tangible the abstract notion of “spectral determinants” yet to come.

18.1 How many ways to get there from here?

In the 3-disk system of example 14.2, the number of admissible trajectories doubles with every iterate. There are $K_n = 3 \cdot 2^n$ distinct itineraries of length n . If disks are too close and a subset of trajectories is pruned, this is only an upper bound and explicit formulas might be hard to discover, but we still might be able to establish a lower exponential bound of the form $K_n \geq Ce^{nh}$. Bounded exponentially by $3e^{n \ln 2} \geq K_n \geq Ce^{nh}$, the number of trajectories must grow exponentially as a function of the itinerary length, with rate given by *topological entropy*,

$$h = \lim_{n \rightarrow \infty} \frac{1}{n} \ln K_n. \quad (18.1)$$

We shall now relate this quantity to the spectrum of the transition matrix, with the growth rate of the number of topologically distinct trajectories given by the leading eigenvalue of the transition matrix.

The transition matrix element $T_{ij} \in \{0, 1\}$ in (17.1) indicates whether the transition from the starting partition j into partition i in one step is allowed or not. Also, the (i, j) element of the transition matrix iterated n times

$$(T^n)_{ij} = \sum_{k_1, k_2, \dots, k_{n-1}} T_{ik_1} T_{k_1 k_2} \dots T_{k_{n-1} j} \quad (18.2)$$

receives a contribution 1 from every admissible sequence of transitions. Thus, $(T^n)_{ij}$ is the number of admissible n symbol itineraries starting with j and ending with i .

exercise 18.1



example 18.1
p. 368

The total number of admissible itineraries of n symbols is

$$K_n = \sum_{ij} (T^n)_{ij} = [1, 1, \dots, 1] T^n \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \quad (18.3)$$

We can also count the number of prime cycles and pruned periodic points, but in order not to break up the flow of the argument, we relegate these pretty results to sect. 18.7. This is recommended reading, if you ever have to compute lots of cycles.

A finite $[N \times N]$ matrix T has eigenvalues $\{\lambda_0, \lambda_1, \dots, \lambda_{m-1}\}$ and (right) eigenvectors $\{\varphi_0, \varphi_1, \dots, \varphi_{m-1}\}$ satisfying $T\varphi_\alpha = \lambda_\alpha \varphi_\alpha$. Expressing the initial vector in (18.3) in this basis (which might be incomplete, with $m \leq N$ eigenvectors),

$$T^n \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = T^n \sum_{\alpha=0}^{m-1} b_\alpha \varphi_\alpha = \sum_{\alpha=0}^{m-1} b_\alpha \lambda_\alpha^n \varphi_\alpha,$$

and contracting with $[1, 1, \dots, 1]$, we obtain

$$K_n = \sum_{\alpha=0}^{m-1} c_\alpha \lambda_\alpha^n.$$

The constants c_α depend on the choice of initial and final partitions. In this example we are sandwiching T^n between the vector $[1, 1, \dots, 1]$ and its transpose, but any other pair of vectors would do, as long as they are not orthogonal to the leading eigenvector φ_0 . In an experiment the vector $[1, 1, \dots, 1]$ would be replaced by a description of the initial state, and the right vector would describe the measurement time n later.

exercise 18.3

Perron theorem states that a Perron-Frobenius matrix has a nondegenerate (isolated) positive real eigenvalue $\lambda_0 > 1$ (with a positive eigenvector) which exceeds the moduli of all other eigenvalues. Therefore as n increases, the sum is dominated by the leading eigenvalue of the transition matrix, $\lambda_0 > |\operatorname{Re} \lambda_\alpha|$, $\alpha = 1, 2, \dots, m-1$, and the topological entropy (18.1) is given by

$$\begin{aligned} h &= \lim_{n \rightarrow \infty} \frac{1}{n} \ln c_0 \lambda_0^n \left[1 + \frac{c_1}{c_0} \left(\frac{\lambda_1}{\lambda_0} \right)^n + \dots \right] \\ &= \ln \lambda_0 + \lim_{n \rightarrow \infty} \left[\frac{\ln c_0}{n} + \frac{1}{n} \frac{c_1}{c_0} \left(\frac{\lambda_1}{\lambda_0} \right)^n + \dots \right] \\ &= \ln \lambda_0, \end{aligned} \tag{18.4}$$

where we have used that $(\lambda_1/\lambda_0)^n$ is small, and the Taylor expansion $\ln(1+x) = x + O(x^2)$.

What have we learned? The transition matrix T is a one-step, *short time* operator, advancing the trajectory from one partition to the next admissible partition. Its eigenvalues describe the rate of growth of the total number of trajectories at the *asymptotic times*. Instead of painstakingly counting $K_1, K_2, K_3, \dots$ and estimating (18.1) from a slope of a log-linear plot, we have the *exact* topological entropy if we can compute the leading eigenvalue of the transition matrix T . This is reminiscent of the way free energy is computed from transfer matrices for 1-dimensional lattice models with finite range interactions. Historically, this analogy, along with statistical mechanics, introduced evolution operator methods into the theory of chaotic systems.

18.2 Topological trace formula

There are two standard ways of computing eigenvalues of a matrix - by evaluating the trace $\operatorname{tr} T^n = \sum \lambda_\alpha^n$, or by evaluating the determinant $\det(1 - zT)$. We start by evaluating the trace of transition matrices. The main lesson will be that the trace receives contributions only from itineraries that return to the initial partition, i.e., periodic orbits.



Consider an M -step memory transition matrix, like the 1-step memory example (17.11). The trace of the transition matrix counts the number of partitions that map into themselves. More generally, each closed walk through n concatenated entries of T contributes to $\text{tr } T^n$, where T^n is the product (18.2) of the matrix entries along the walk. Each step in such a walk shifts the symbolic string by one symbol; the trace ensures that the walk closes on a periodic string c . Define t_c to be the product of matrix elements along a cycle c , each term being multiplied by a book keeping variable z .

The lower case ‘ t ’ indicates that the quantity t_c is a ‘local trace’ associated with the particular, ‘local’ walk c , in the sense that the trace of T^n is a sum over such quantities. In the following chapters, t_c takes on a continuum of values, so for the remainder of this chapter we use the ‘ t_c ’ notation rather than to the 0 or z^n values specific to the counting problem. The bookkeeping variable z is for the moment just that, a bookkeeping device, but in the chapters to come it will assume a much deeper role as the Laplace transform variable dual to the discrete time n (see the discussion following (21.8), (21.10)), just as the energy is the variable dual to time in quantum mechanics (this is made explicit by the relation between continuous and discrete time cases (21.20), and, deeper still, by the form of the semiclassical zeta function (39.12)).

The quantity $z^n \text{tr } T^n$ is then the sum of t_c for all cycles of period n . The t_c = (product of matrix elements along cycle c) is manifestly cyclically invariant, $t_{100} = t_{010} = t_{001}$, so a prime cycle p of period n_p contributes n_p times, once for each periodic point along its orbit. For the purpose of periodic orbit counting (remember the definition of the transition matrix in (17.1)), the local trace takes values

$$t_p = \begin{cases} z^{n_p} & \text{if } p \text{ is an admissible cycle} \\ 0 & \text{otherwise,} \end{cases} \quad (18.5)$$

i.e., (setting $z = 1$) the local trace is $t_p = 1$ if the cycle is admissible, and $t_p = 0$ otherwise.

 example 18.2
p. 368

Hence $\text{tr } T^n = N_n$ counts the number of *admissible periodic points* of period n . The n th order trace (18.29) picks up contributions from all repeats of prime cycles, with each cycle contributing n_p periodic points, so N_n , the total number of periodic points of period n is given by

example 18.1

$$z^n N_n = z^n \text{tr } T^n = \sum_{n_p | n} n_p t_p^{n/n_p} = \sum_p n_p \sum_{r=1}^{\infty} \delta_{n_p r, n} t_p^r. \quad (18.6)$$

Here $m|n$ means that m is a divisor of n . An example is the periodic orbit counting in table 18.2.

remark 18.1

In order to get rid of the awkward divisibility constraint $n = n_p r$ in the above sum, we introduce the generating function for numbers of periodic points

remark A22.1

Table 18.1: Prime cycles for the binary symbolic dynamics up to period 9. The numbers of prime cycles are given in table 18.3.

n_p	p	n_p	p	n_p	p	n_p	p	n_p	p	n_p	p
1	0	7	0001001	8	00001111	9	000001101	9	001001111		
	1		0000111		00010111		000010011		001010111		
2	01		0001011		00011011		000010101		001011011		
3	001		0001101		00011101		000011001		001011101		
	011		0010011		00100111		000100011		001100111		
4	0001		0010101		00101011		000100101		001101011		
	0011		0001111		00101101		000101001		001101101		
	0111		0010111		00110101		000001111		001110101		
5	00001		0011011		00011111		000010111		010101011		
	00011		0011101		00101111		000011011		000111111		
	00101		0101011		00110111		000011101		001011111		
	00111		0011111		00111011		000100111		001101111		
	01011		0101111		00111101		000101011		001110111		
	01111		0110111		01010111		000101101		001111011		
6	000001		0111111		01011011		000110011		001111101		
	000011	8	00000001		00111111		000110101		010101111		
	000101		00000011		01011111		000111001		010110111		
	000111		00000101		01101111		001001011		010111011		
	001011		00001001		01111111		001001101		001111111		
	001101		00000111	9	000000001		001010011		010111111		
	001111		00001011		000000011		001010101		011011111		
	010111		00001101		000000101		000011111		011101111		
	011111		00010011		000001001		000101111		011111111		
7	0000001		00010101		000010001		000110111				
	0000011		00011001		000000111		000111011				
	0000101		00100101		000001011		000111101				

Table 18.2: The total number N_n of periodic points of period n , expressed in terms of prime cycles (18.6), for binary symbolic dynamics. The number of contributing prime cycles illustrates the preponderance of long prime cycles of period n over the repeats of shorter cycles of periods n_p , where $n = rn_p$. Further enumerations of binary prime cycles are given in tables 18.1 and 18.3. (L. Rondoni)

n	N_n	# of prime cycles of period n_p									
		1	2	3	4	5	6	7	8	9	10
1	2	2									
2	4	2	1								
3	8	2		2							
4	16	2	1		3						
5	32	2				6					
6	64	2	1	2			9				
7	128	2						18			
8	256	2	1		3				30		
9	512	2		2						56	
10	1024	2	1			6					99

$$\sum_{n=1}^{\infty} z^n N_n = \text{tr} \frac{zT}{1 - zT} . \quad (18.7)$$

The right hand side is the geometric series sum of $N_n = \text{tr} T^n$. Substituting (18.6) into the left hand side, and replacing the right hand side by the eigenvalue sum $\text{tr} T^n = \sum \lambda_{\alpha}^n$, we obtain our first example of a trace formula, the *topological trace formula*

$$\sum_{\alpha=0} \frac{z \lambda_{\alpha}}{1 - z \lambda_{\alpha}} = \sum_p \frac{n_p t_p}{1 - t_p} . \quad (18.8)$$

A trace formula relates the spectrum of eigenvalues of an operator - here the transition matrix - to the spectrum of periodic orbits of a dynamical system. A statement of duality exists between short-time, local information and global averages. In this case, local information refers to the next admissible symbol in a symbol sequence, and global averages represent the mean rate of growth of the number of cycles with increasing cycle period.

The z^n sum in (18.7) is a discrete version of the Laplace transform (see sect. 21.1.2), and the resolvent on the left hand side is the antecedent of the more sophisticated trace formulas (21.9), (21.19), (21.24) and the Gutzwiller trace formula (39.3) of semiclassical quantum mechanics. We shall now use this result to compute the spectral determinant of the transition matrix.

18.3 Determinant of a graph

Our next task is to determine the zeros of the *spectral determinant* of an $[m \times m]$ transition matrix

$$\det(1 - zT) = \prod_{\alpha=0}^{m-1} (1 - z \lambda_{\alpha}) . \quad (18.9)$$

We could now proceed to diagonalize T on a computer and get this over with. It pays, however, to dissect $\det(1 - zT)$ with some care; understanding this computation in detail will be the key to understanding the cycle expansion computations of chapter 23 for arbitrary dynamical averages. For T a finite matrix (18.9) is just the characteristic polynomial for T . However, we shall be able to compute this object even when the dimension of T and other such operators becomes infinite, and for that reason we prefer to refer to (18.9) loosely as the “spectral determinant.”

There are various definitions of the determinant of a matrix; we will view the determinant as a sum over all possible permutation cycles composed of the traces $\text{tr} T^k$, in the spirit of the determinant–trace relation (1.16):

$$\begin{aligned} \det(1 - zT) &= \exp(\text{tr} \ln(1 - zT)) = \exp\left(-\sum_{n=1} \frac{z^n}{n} \text{tr} T^n\right) \\ &= 1 - z \text{tr} T - \frac{z^2}{2} ((\text{tr} T)^2 - \text{tr} T^2) - \dots \end{aligned} \quad (18.10)$$



exercise 4.1
example A20.1

This is sometimes called a *cumulant* expansion. Formally, the right hand is a Taylor series in z about $z = 0$. If T is an $[m \times m]$ finite matrix, then the characteristic polynomial is at most of order m . In that case the coefficients of z^n must vanish *exactly* for $n > m$.

We now proceed to relate the determinant in (18.10) to the corresponding transition graph of chapter 17. Toward this end, we start with the usual textbook expression for a determinant as the sum of products of all permutations

$$\det M = \sum_{\{\pi\}} (-1)^\pi M_{1,\pi_1} M_{2,\pi_2} \cdots M_{m,\pi_m}, \quad (18.11)$$

where $M = 1 - zT$ is a $[m \times m]$ matrix, $\{\pi\}$ denotes the set of permutations of m symbols, π_k is the permutation π applied to k , and $(-1)^\pi = \pm 1$ is the parity of permutation π . The right hand side of (18.11) yields a polynomial in T of order m in z . A contribution of order n in z picks up $m - n$ unit factors along the diagonal, and the remaining matrix elements yield

$$(-z)^n (-1)^\pi T_{s_1 \pi s_1} \cdots T_{s_n \pi s_n}, \quad (18.12)$$

where π is the permutation of the subset of n distinct symbols $s_1 \cdots s_n$ indexing T matrix elements. As in (18.29), we refer to any combination $t_c = T_{s_1 s_k} T_{s_3 s_2} \cdots T_{s_2 s_1}$, for a given itinerary $c = s_1 s_2 \cdots s_k$, as the *local trace* associated with a closed loop c on the transition graph. Each term of the form (18.12) may be factored in terms of local traces $t_{c_1} t_{c_2} \cdots t_{c_k}$, i.e., loops on the transition graph. These loops are non-intersecting, as each node may only be reached by *one* link. They are indeed loops, and if a node is reached by a link, it has to be the starting point of another *single* link, as each s_j must appear exactly *once* as a row and column index.

The general structure is clear, but a little more thinking is required to determine the sign of a generic contribution. We consider only the case of loops of length 1 and 2, and we leave the reader with the task of generalizing the result by induction. Consider first a term in which only loops of unit length appear in (18.12), i.e., only the diagonal elements of T are picked up. We have $k = m$ loops and an even permutation π , so the sign is given by $(-1)^k$, where k is the number of loops. Now take the case in which we have i single loops and j loops of length $n = 2j + i$. The parity of the permutation gives $(-1)^j$ and the first factor in (18.12) gives $(-1)^n = (-1)^{2j+i}$. So once again these terms combine to $(-1)^k$, where $k = i + j$ is the number of loops. Let f be the maximal number of non-intersecting loops. We may summarize our findings as follows:

exercise 18.4

The characteristic polynomial of a transition matrix is given by the sum of all possible partitions π of the corresponding transition graph into products of k non-intersecting loops, with each loop trace t_p carrying a minus sign:

$$\det(1 - zT) = \sum_{k=0}^f \sum_{\pi}' (-1)^k t_{p_1} \cdots t_{p_k} \quad (18.13)$$



Any self-intersecting loop is *shadowed* by a product of two loops that share the intersection point. In the case at hand, both the long loop t_{ab} and its shadow t_{at_b} carry the same weight $z^{n_a+n_b}$. The cancelation is exact, and the loop expansion (18.13) is finite. In the case where the local traces count prime cycles (18.5), $t_p = 0$ or z^n , we refer to $\det(1 - zT)$ as the *topological polynomial*.

We refer to the set of all non-self-intersecting loops $\{t_{p_1}, t_{p_2}, \dots, t_{p_f}\}$ as *fundamental cycles* (for an explicit example, see the loop expansion of example 18.6). This is not a very good definition, as transition graphs are not unique. The most we know is that for a given finite-grammar language, there exist transition graph(s) with a minimal number of loops. Regardless of how cleverly a transition graph is constructed, it is always true that for any finite transition graph the number of fundamental cycles f is finite. If the graph has m nodes, no fundamental cycle has a period longer than m , since any longer cycle is self-intersecting by necessity.

The above loop expansion of a determinant in terms of traces is most easily grasped by working through a few examples. The complete binary dynamics transition graph of figure 17.5 is a little bit too simple, but let us start humbly and consider it anyway.



example 18.3
p. 369

Similarly, for the complete symbolic dynamics of N symbols the transition graph has one node and N links, yielding

$$\det(1 - zT) = 1 - Nz, \quad (18.14)$$

which gives the topological entropy $h = \ln N$.



example 18.4
p. 369



example 18.5
p. 369



example 18.6
p. 370

18.4 Topological zeta function

What happens if there is no finite-memory transition matrix, meaning that the transition graph is infinite? What if we are never sure that looking further into the future reveals any further forbidden blocks? There is still a way to define the determinant, and this idea is central to the whole treatise. The determinant is then defined by its *cumulant* expansion (18.10)

exercise 4.1

$$\det(1 - zT) = 1 - \sum_{n=1}^{\infty} \hat{c}_n z^n. \quad (18.15)$$



example 18.7
p. 371

For finite dimensional matrices the expansion is a finite polynomial, and (18.15) is an identity equation; however, for infinite dimensional operators the cumulant expansion coefficients $\hat{c}_n$ define the determinant.

Let us now evaluate the determinant in terms of traces for an arbitrary transition matrix. In order to obtain an expression for the spectral determinant (18.9) in terms of cycles, substitute (18.6) into (18.15) and sum over the repeats of prime cycles using $\ln(1 - x) = -\sum_r x^r/r$,

$$\begin{aligned} \det(1 - zT) &= \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{t_p^r}{r}\right) = \exp\left(\sum_p \ln(1 - t_p)\right) \\ \prod_{\alpha} (1 - z\lambda_{\alpha}) &= \prod_p (1 - t_p), \end{aligned} \quad (18.16)$$

where for the topological entropy the weight assigned to a prime cycle p of period n_p is $t_p = z^{n_p}$ if the cycle is admissible, or $t_p = 0$ if it is pruned. This determinant is called the *topological* or *Artin-Mazur* zeta function, conventionally denoted by



$$1/\zeta_{\text{top}}(z) = \prod_p (1 - z^{n_p}) = 1 - \sum_{n=1} \hat{c}_n z^n. \quad (18.17)$$

Counting cycles amounts to giving each admissible prime cycle p weight $t_p = z^{n_p}$ and expanding the Euler product (18.17) as a power series in z . The number of prime cycles p is infinite, but if T is an $[m \times m]$ finite matrix, then the number of roots λ_{α} is at most m . The characteristic polynomial is at most of order m , and the coefficients of z^n vanish for $n > m$. Since the precise expression for the coefficients $\hat{c}_n$, in terms of local traces t_p , is more general than the current application to counting, we postpone its derivation to chapter 23.

The topological entropy h can now be determined from the leading zero $z = e^{-h}$ of the topological zeta function. For a finite $[m \times m]$ transition matrix, the number of terms in the characteristic equation (18.13) is finite, and we refer to this expansion as the *topological polynomial* of order $\leq m$. The utility of defining the determinant by its cumulant expansion is that it works even when the partition is infinite, $m \rightarrow \infty$. An example is given in sect. 18.5, and many more examples will be given later on.



example 18.13
p. 373



example 18.14
p. 374



fast track:
sect. 18.5, p. 358

18.4.1 Topological zeta function for flows



We now apply the method that we shall use in deriving (21.19) to the problem of deriving the topological zeta functions for flows. The time-weighted density of prime cycles of period t is

$$\Gamma(t) = \sum_p \sum_{r=1}^{\infty} T_p \delta(t - rT_p). \quad (18.18)$$

The Laplace transform smooths the sum over Dirac delta spikes (see (21.18)) and yields the *topological trace formula*

$$\sum_p \sum_{r=1}^{\infty} T_p \int_{0+}^{\infty} dt e^{-st} \delta(t - rT_p) = \sum_p T_p \sum_{r=1}^{\infty} e^{-sT_p r} \quad (18.19)$$

and the *topological zeta function* for flows:

$$1/\zeta_{\text{top}}(s) = \prod_p (1 - e^{-sT_p}), \quad (18.20)$$

related to the trace formula by

$$\sum_p T_p \sum_{r=1}^{\infty} e^{-sT_p r} = -\frac{\partial}{\partial s} \ln 1/\zeta_{\text{top}}(s).$$

This is the continuous time version of the discrete time topological zeta function (18.17) for maps; its leading zero $s = -h$ yields the topological entropy for a flow.

18.5 Topological zeta function for an infinite partition

(K.T. Hansen and P. Cvitanović)



To understand the need for topological zeta function (18.15), we turn to a dynamical system with (as far as we know - there is no proof) an infinite partition or with an infinity of ever-longer pruning rules. Consider the 1-dimensional quadratic map (14.20)

$$f(x) = Ax(1 - x), \quad A = 3.8.$$

Numerically the kneading sequence (the itinerary of the critical point $x = 1/2$ (14.5)) is

exercise 18.20

$$K = 101101111011011110101111011110\dots$$

where the symbolic dynamics is defined by the partition of figure 14.9. How this kneading sequence is converted into a series of pruning rules is a dark art. For the moment, it suffices to state the result in order to give you a feeling for what a

Figure 18.1: The logarithm $\ln |z_0^{(n)} - z_0|$ of the difference between the leading zero of the n th polynomial approximation to topological zeta function and our best estimate (18.23), as a function of order of the polynomial n (the topological zeta function evaluated for the closest value of A to $A = 3.8$ for which the quadratic map has a stable cycle of period n). (from K.T. Hansen [12])

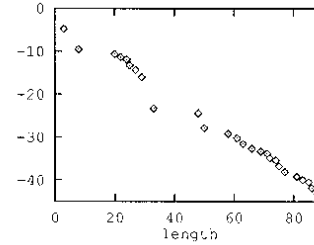
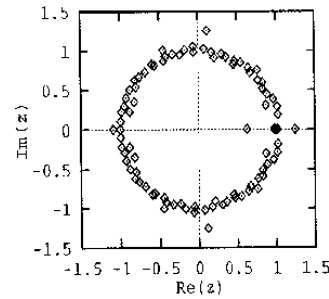


Figure 18.2: The 90 zeroes of the topological zeta function for the quadratic map for $A = 3.8$ approximated by the nearest topological zeta function with a stable cycle of period 90. (from K.T. Hansen [12])



“typical” infinite partition topological zeta function looks like. For example, approximating the dynamics by a transition graph corresponding to a repeller of the period 29 attractive cycle close to the $A = 3.8$ strange attractor yields a transition graph with 29 nodes and the characteristic polynomial

$$\begin{aligned} 1/\zeta_{\text{top}}^{(29)} = & 1 - z^1 - z^2 + z^3 - z^4 - z^5 + z^6 - z^7 + z^8 - z^9 - z^{10} \\ & + z^{11} - z^{12} - z^{13} + z^{14} - z^{15} + z^{16} - z^{17} - z^{18} + z^{19} + z^{20} \\ & - z^{21} + z^{22} - z^{23} + z^{24} + z^{25} - z^{26} + z^{27} - z^{28}. \end{aligned} \quad (18.21)$$

The smallest real root of this approximate topological zeta function is

$$z = 0.62616120 \dots \quad (18.22)$$

Constructing finite transition graphs of increasing length corresponding to $A \rightarrow 3.8$, we find polynomials with better and better estimates for the topological entropy. For the closest stable period 90 orbit, we obtain our best estimate of the topological entropy of the repeller:

$$h = -\ln 0.62616130424685 \dots = 0.46814726655867 \dots \quad (18.23)$$

Figure 18.1 illustrates the convergence of the truncation approximations to the topological zeta function. In this graph, the logarithm of the difference between the zero of a polynomial and our best estimate (18.23) is plotted as a function of the period of the stable periodic orbit. The error of the estimate (18.22) is expected to be of order $z^{29} \approx e^{-14}$, because going from period 28 to a longer truncation typically yields combinations of loops with 29 and more nodes giving terms $\pm z^{29}$ and of higher order in the polynomial. Hence the convergence is exponential, with an exponent of $-0.47 = -h$, the topological entropy itself. In figure 18.2

we plot the zeroes of the polynomial approximation to the topological zeta function obtained by accounting for all forbidden strings of length 90 or less. The leading zero giving the topological entropy is the point closest to the origin. Most of the other zeroes are close to the unit circle; we conclude that for infinite state space partitions the topological zeta function has a unit circle as the radius of convergence. The convergence is controlled by the ratio of the leading eigenvalue to the next-to-leading eigenvalue. In this case, the ratio is indeed $\lambda_1/\lambda_0 = 1/e^h = e^{-h}$.

18.6 Shadowing

The topological zeta function is a pretty function, but the infinite product (18.16) should make you pause. For finite transition matrices, the left hand side is a determinant of a finite matrix and, therefore, a finite polynomial. So why is the right hand side an infinite product over the infinitely many prime periodic orbits of all periods?

The way in which this infinite product rearranges itself into a finite polynomial is instructive and crucial for all that follows. You can already take a peek at the full cycle expansion (23.8) of chapter 23; all cycles beyond the fundamental t_0 and t_1 appear in the shadowing combinations such as

$$t_{s_1 s_2 \dots s_n} - t_{s_1 s_2 \dots s_m} t_{s_{m+1} \dots s_n}.$$

For subshifts of finite type such shadowing combinations cancel *exactly*, if we are counting cycles as we do in (18.30) and (18.37), or if the dynamics is piecewise linear, as in exercise A22.3. As we argue in sect. 1.5.4, for nice hyperbolic flows whose symbolic dynamics is a subshift of finite type, the shadowing combinations *almost* cancel. Additionally, the spectral determinant is dominated by the fundamental cycles from (18.13), with longer cycles contributing only small “curvature” corrections.

These exact or nearly exact cancelations depend on the flow being smooth and the symbolic dynamics being a subshift of finite type. If the dynamics requires an infinite state space partition, with pruning rules for blocks of increasing length, most of the shadowing combinations still cancel. However, a few blocks are forbidden and do not lead to a finite radius of convergence for the spectral determinant, as depicted in figure 18.2.

One striking aspect of the pruned cycle expansion (18.21) compared to the trace formulas such as (18.7) is that coefficients are not growing exponentially - indeed they all remain of order 1. Instead of having a radius of convergence e^{-h} , as in the example at hand, the topological zeta function has the unit circle as the radius of convergence. In other words, exponentiating the spectral problem from a trace formula to a spectral determinant as in (18.15) increases the *analyticity domain*. The pole in the trace (18.8) at $z = e^{-h}$ is promoted to a smooth zero of the spectral determinant with a larger radius of convergence.

This sensitive dependence of spectral determinants on whether or not the symbolic dynamics is a subshift of finite type is bad news. If the system is generic and not structurally stable (see sect. 15.2), a smooth parameter variation is in no sense a smooth variation of topological dynamics. Infinities of periodic orbits are created or destroyed, and transition graphs go from being finite to infinite and back. That will imply that the global averages that we intend to compute are generically not differentiable functions of the system parameters, and averaging over families of dynamical systems can be a highly nontrivial enterprise. A simple illustration is the parameter dependence of the diffusion constant computed in a remark in chapter 24.

You might be asking, “What is wrong with computing the entropy from (18.1)? Does all this theory buy us anything?” An answer, if we count K_n level by level, we ignore the self-similarity of the pruned tree - examine for example figure 17.6, or the cycle expansion of (18.38) - and the finite estimates of $h_n = \ln K_n/n$ converge nonuniformly to h , and on top of that with a slow rate of convergence, $|h - h_n| \approx O(1/n)$ as in (18.4). By construction, the determinant (18.9) is much smarter. It encodes the self-similarity of the dynamics and yields the asymptotic value of h , with no need for any finite n extrapolations.



fast track:
sect. 19, p. 381

18.7 Counting cycles

Chaos is merely order waiting to be deciphered
— José Saramago, *The Double*



In what follows, we shall occasionally need to compute all cycles up to topological period n , so it is important to know their exact number. The formulas are fun to derive, but a bit technical for a plumber on the street, and they are probably best skipped on the first reading.

18.7.1 Counting periodic points

The number of periodic points of period n is denoted N_n . It can be computed from (18.15) and (18.7) as a logarithmic derivative of the topological zeta function

$$\begin{aligned} \sum_{n=1} N_n z^n &= \operatorname{tr} \left(-z \frac{d}{dz} \ln(1 - zT) \right) = -z \frac{d}{dz} \ln \det (1 - zT) \\ &= \frac{-z \frac{d}{dz} (1/\zeta_{\text{top}})}{1/\zeta_{\text{top}}}. \end{aligned} \quad (18.24)$$

Table 18.3: Number of prime cycles for various alphabets and grammars up to period 10. The first column gives the cycle period, the second gives the formula (18.26) for the number of prime cycles for complete N -symbol dynamics, and columns three through five give the numbers of prime cycles for $N = 2, 3$ and 4.

n	$M_n(N)$	$M_n(2)$	$M_n(3)$	$M_n(4)$
1	N	2	3	4
2	$N(N-1)/2$	1	3	6
3	$N(N^2-1)/3$	2	8	20
4	$N^2(N^2-1)/4$	3	18	60
5	$(N^5-N)/5$	6	48	204
6	$(N^6-N^3-N^2+N)/6$	9	116	670
7	$(N^7-N)/7$	18	312	2340
8	$N^4(N^4-1)/8$	30	810	8160
9	$N^3(N^6-1)/9$	56	2184	29120
10	$(N^{10}-N^5-N^2+N)/10$	99	5880	104754

Observe that the trace formula (18.8) diverges at $z \rightarrow e^{-h}$, because the denominator has a simple zero there.



example 18.8
p. 371



example 18.9
p. 372

18.7.2 Counting prime cycles

Having calculated the number of periodic points, our next objective is to evaluate the number of *prime* cycles M_n for a dynamical system whose symbolic dynamics is built from N symbols. The problem of finding M_n is classical in combinatorics (counting necklaces made out of n beads of N different kinds) and is easily solved. There are N^n possible distinct strings of length n composed of N letters. These N^n strings include all M_d prime d -cycles whose period d equals or divides n . A prime cycle is a non-repeating symbol string. For example, $p = 011 = \overline{101} = \overline{110} = \dots 011011 \dots$ is prime, but $\overline{0101} = 010101 \dots = \overline{01}$ is not. A prime d -cycle contributes d strings to the sum of all possible strings, one for each cyclic permutation. The total number of possible periodic symbol sequences of period n is therefore related to the number of prime cycles by

$$N_n = \sum_{d|n} dM_d, \quad (18.25)$$

where N_n equals $\text{tr } T^n$. The number of prime cycles can be computed recursively

$$M_n = \frac{1}{n} \left(N_n - \sum_{d|n, d < n} dM_d \right),$$

or by the *Möbius inversion formula*

$$M_n = n^{-1} \sum_{d|n} \mu\left(\frac{n}{d}\right) N_d, \quad (18.26)$$

exercise 18.10

Table 18.4: List of 3-disk prime cycles up to period 10. Here n is the cycle period, M_n is the number of prime cycles, N_n is the number of periodic points, and S_n is the number of distinct prime cycles under D_3 symmetry (see chapter 25 for further details). Column 3 also indicates the splitting of N_n into contributions from orbits of periods that divide n . The prefactors in the fifth column indicate the degeneracy m_p of the cycle; for example, 3·12 stands for the three prime cycles $\overline{12}$, $\overline{13}$ and $\overline{23}$ related by $2\pi/3$ rotations. Among symmetry-related cycles, a representative $\hat{p}$ which is lexically lowest is listed. The cycles of period 9 grouped with parentheses are related by time reversal symmetry, but not by any D_3 transformation.

n	M_n	N_n	S_n	$m_p \cdot \hat{p}$
1	0	0	0	
2	3	6=3·2	1	3·12
3	2	6=2·3	1	2·123
4	3	18=3·2+3·4	1	3·1213
5	6	30=6·5	1	6·12123
6	9	66=3·2+2·3+9·6	2	6·121213 + 3·121323
7	18	126=18·7	3	6·1212123 + 6·1212313 + 6·1213123
8	30	258=3·2+3·4+30·8	6	6·12121213 + 3·12121313 + 6·12121323 + 6·12123123 + 6·12123213 + 3·12132123
9	56	510=2·3+56·9	10	6·121212123 + 6·(121212313 + 121212323) + 6·(121213123 + 121213213) + 6·121231323 + 6·(121231213 + 121232123) + 2·121232313 + 6·121321323
10	99	1022	18	

where the Möbius function is 0 if n has a squared prime factor, and is $(-1)^k$ if n is a product of k distinct primes. Thus, $\mu(1) = 1$, $\mu(n) = 0$, if n has a squared factor, and $\mu(p_1 p_2 \dots p_k) = (-1)^k$, if all prime factors are different.

We list the number of prime cycles up to period 10 for 2-, 3- and 4-letter complete symbolic dynamics in table 18.3, obtained by Möbius inversion (18.26).

exercise 18.11



example 18.10
p. 372



example 18.11
p. 372



example 18.12
p. 372

Résumé



The main result of this chapter is the cycle expansion (18.17) of the topological zeta function (i.e., the spectral determinant of the transition matrix):

$$1/\zeta_{\text{top}}(z) = 1 - \sum_{k=1} \hat{c}_k z^k.$$

For subshifts of finite type, the transition matrix is finite, and the topological zeta function is a finite polynomial evaluated by the loop expansion (18.13) of $\det(1 - zT)$. For infinite grammars the topological zeta function is defined by its cycle expansion. The topological entropy h is given by the leading zero $z = e^{-h}$. This expression for the entropy is *exact*; in contrast to the initial definition (18.1), no $n \rightarrow \infty$ extrapolations of $\ln K_n/n$ are required.

Table 18.5: The 4-disk prime cycles up to period 8. The symbol definitions are the same as those shown in table 18.4. Orbits related by the time reversal symmetry (but no D_4 symmetry) already appear at cycle period 5. Cycles of period 7 and 8 have been omitted.

n	M_n	N_n	S_n	$m_p \cdot \hat{p}$
1	0	0	0	
2	6	$12=6 \cdot 2$	2	$4 \cdot 12 + 2 \cdot 13$
3	8	$24=8 \cdot 3$	1	$8 \cdot 123$
4	18	$84=6 \cdot 2 + 18 \cdot 4$	4	$8 \cdot 1213 + 4 \cdot 1214 + 2 \cdot 1234 + 4 \cdot 1243$
5	48	$240=48 \cdot 5$	6	$8 \cdot (12123 + 12124) + 8 \cdot 12313$ $+ 8 \cdot (12134 + 12143) + 8 \cdot 12413$
6	116	$732=6 \cdot 2 + 8 \cdot 3 + 116 \cdot 6$	17	$8 \cdot 121213 + 8 \cdot 121214 + 8 \cdot 121234$ $+ 8 \cdot 121243 + 8 \cdot 121313 + 8 \cdot 121314$ $+ 4 \cdot 121323 + 8 \cdot (121324 + 121423)$ $+ 4 \cdot 121343 + 8 \cdot 121424 + 4 \cdot 121434$ $+ 8 \cdot 123124 + 8 \cdot 123134 + 4 \cdot 123143$ $+ 4 \cdot 124213 + 8 \cdot 124243$
7	312	2184	39	
8	810	6564	108	

What have we accomplished? We have related the number of topologically distinct paths from one state space region to another region to the leading eigenvalue of the transition matrix T . The spectrum of T is given by topological zeta function, a certain sum over traces $\text{tr } T^n$, and in this way the periodic orbit theory has entered the arena through the trace formula (18.8), already at the level of the topological dynamics.

The main lesson of learning how to count well, a lesson that will be constantly reaffirmed, is that while trace formulas are a conceptually essential step in deriving and understanding periodic orbit theory, the spectral determinant is the right object to use in actual computations. Instead of summing all of the exponentially many periodic points required by trace formulas at each level of truncation, spectral determinants incorporate only the small incremental corrections to what is already known - and that makes them a more powerful tool for computations.

Contrary to claims one all too often encounters in the literature, “exponential proliferation of trajectories” is not the problem. The limiting factor for the convergence of cycle expansions is the proliferation of the grammar rules or “algorithmic complexity,” as illustrated by sect. 18.5 and figure 18.2, in particular. Nice, finite grammar leads to nice, discrete spectrum; infinite grammar leads to analytical walls in the complex spectral plane.

Historically, these topological zeta functions were the inspiration for applying the transfer matrix methods of statistical mechanics to the problem of computation of dynamical averages for chaotic flows. The key result was the dynamical zeta function to be derived in chapter 21, a weighted generalization of the topological zeta function.

Commentary

Remark 18.1. Generating functions, Z-transforms. The method of generating functions was introduced by de Moivre in 1730. Euler used it, for example in 1748, to partition integers [14]. Laplace named it ‘generating function’ [20]. In 1947 Hurewicz [15] used it to solve linear, constant-coefficient difference equations. In 1952 Ragazzini and Zadeh [21] renamed it the ‘Z-transforms’, as generating functions are but a form of Laplace transforms. The theory is pedagogically explained by Elaydi [10], including a table of common Z-transform pairs, in analogy with the familiar Laplace transform tables.

Remark 18.2. Artin-Mazur zeta functions. Motivated by A. Weil’s zeta function for the Frobenius map [25], Artin and Mazur [2] introduced the zeta function (18.17) that counts periodic points for diffeomorphisms (see also ref. [19] for their evaluation for maps of the interval). Smale [24] conjectured rationality of the zeta functions for Axiom A diffeomorphisms, later proved by Guckenheimer [11] and Manning [18]. Every subshift of finite type has a rational zeta function. However, most subshifts have irrational zeta functions [5]. See remark 22.4 on page 441 for more zeta function history.

Remark 18.3. “Entropy.” The ease with which the topological entropy can be motivated obscures the fact that our construction does not lead to an invariant characterization of the dynamics, as the choice of symbolic dynamics is largely arbitrary. The same caveat applies to other entropies. In order to obtain invariant characterizations we will have to work harder. Mathematicians like to define the (impossible to evaluate) supremum over all possible partitions. The key point that eliminates the need for such searches is the existence of *generators*, i.e., partitions that are able to probe the whole state space on arbitrarily small scales. A generator is a finite partition $\mathcal{M} = \{\mathcal{M}_1 \dots \mathcal{M}_N\}$ with certain properties. For example, consider the partition built upon all possible intersections of sets $f^n(\mathcal{M}_i)$, where f is dynamical evolution and n takes all possible integer values (positive as well as negative), then the closure of such a partition coincides with the ‘algebra of all measurable sets.’ For a thorough (and readable) discussion of generators and how they allow a computation of the Kolmogorov entropy, see Arnol’d & Avez [1].

Remark 18.4. Perron-Frobenius matrices. For a proof of the Perron theorem on the leading eigenvalue see ref. [16]. Appendix A4.1 of Zinn-Justin monograph [26] offers a clear discussion of the spectrum of the transition (or Perron-Frobenius) matrix.

Remark 18.5. Determinant of a graph. Many textbooks offer derivations of the loop expansions of characteristic polynomials for transition matrices and their transition graphs, see for example refs. [7, 13, 23].

Remark 18.6. Ordering periodic orbit expansions. In sect. 23.7 we will introduce an alternative way of hierarchically organizing cumulant expansions, in which the order is dictated by stability rather than cycle period: such a procedure may be better suited to perform computations when the symbolic dynamics is not well understood.

Remark 18.7. T is not trace class. Note to the erudite reader: the transition matrix T (in the infinite partition limit (18.15)) is *not* trace class. Still the trace is well defined in the $n \rightarrow \infty$ limit.

Remark 18.8. Counting prime cycles. Sequences of objects that are identical up to cyclic permutations are called (Pólya) necklaces [22]. If the number of objects they consist of is two, they are called binary necklaces. Duval [3, 9] has an efficient algorithm for generating Lyndon words [6, 17] (non-periodic necklaces, i.e., prime cycle itineraries). For a discussion of Möbius function, see Dummit and Foote [8].

References

- [1] V. I. Arnol'd and A. Avez, *Ergodic Problems of Classical Mechanics* (Addison-Wesley, Redwood City, 1989).
- [2] M. Artin and B. Mazur, “On periodic points”, *Ann. Math.* **81**, 82–99 (1965).
- [3] J. Berstel and M. Pocchiola, “Average cost of Duval’s algorithm for generating Lyndon words”, *Theor. Comput. Sci.* **132**, 415–425 (1994).
- [4] N. Bird and F. Vivaldi, “Periodic orbits of the sawtooth maps”, *Physica D* **30**, 164–176 (1988).
- [5] R. Bowen and O. Lanford, Zeta functions of restrictions of the shift transformation, in *Global Analysis (Proc. Sympos. Pure Math., Berkeley, CA, 1968)*, Vol. 1, edited by S.-S. Chern and S. Smale (1970), pp. 43–50.
- [6] M. Chemillier, “Periodic musical sequences and Lyndon words”, *Soft Comput.* **8**, 611–616 (2004).
- [7] D. M. Cvektović, M. Doob, and H. Sachs, *Spectra of Graphs* (Academic, New York, 1980).
- [8] D. S. Dummit and R. M. Foote, *Abstract Algebra* (Wiley, 2003).
- [9] J.-P. Duval, “Generation d’une section des classes de conjugaison et arbre des mots de Lyndon de longueur borné”, *Theor. Comput. Sci.* **60**, 255–283 (1988).
- [10] S. Elaydi, *An Introduction to Difference Equations* (Springer, Berlin, 2005).
- [11] J. Guckenheimer, “Axiom A + NO cycles $\rightarrow \zeta_f(t)$ rational”, *Bull. Amer. Math. Soc.* **76**, 592–594 (1970).
- [12] K. T. Hansen, *Symbolic Dynamics in Chaotic systems*, PhD thesis (Univ. of Oslo, Oslo, Norway, 1993).
- [13] J. E. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages and Computation* (Addison-Wesley, Reading MA, 1979).
- [14] B. Hopkins and R. Wilson, “Euler’s science of combinations”, in *Leonhard Euler: Life, Work and Legacy*, edited by R. E. Bradley and C. E. Sandifer (Elsevier, 2007), pp. 395–408.
- [15] W. Hurewicz, “Filters and servo systems with pulsed data”, in *Theory of Servomechanisms*, edited by H. M. James, N. B. Nichols, and R. S. Phillips (McGraw-Hill, New York, 1947), pp. 231–261.
- [16] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems* (Cambridge Univ. Press, Cambridge UK, 1995).

- [17] R. C. Lyndon, *Groups and Geometry* (Cambridge Univ. Press, Cambridge UK, 1985).
- [18] A. Manning, “Axiom A diffeomorphisms have rational zeta function”, *Bull. London Math. Soc.* **3**, 215–220 (1971).
- [19] J. Milnor and W. Thurston, “On iterated maps of the interval”, in *Dynamical Systems*, edited by J. C. Alexander (Springer, New York, 1988), pp. 465–563.
- [20] G. Polya, *Mathematics and Plausible Reasoning, Vol. 1 Induction and Analogy in Mathematics* (Princeton Univ. Press, 1954).
- [21] J. R. Ragazzini and L. A. Zadeh, “The analysis of sampled-data systems”, *Trans. AIEE* **71**, 225–234 (1952).
- [22] J. Riordan, *An Introduction to Combinatorial Analysis* (Wiley, New York, 1958).
- [23] A. Salomaa, *Formal Languages* (Academic, San Diego, 1973).
- [24] S. Smale, “Differentiable dynamical systems”, *Bull. Amer. Math. Soc.* **73**, 747–817 (1967).
- [25] A. Weil, “Numbers of solutions of equations in finite fields”, *Bull. Amer. Math. Soc.* **55**, 497–508 (1949).
- [26] J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena* (Oxford Univ. Press, Oxford, 1989).

18.8 Examples

Example 18.1. 3-disk itinerary counting. Consider the 3-state partition transition matrix

$$T = \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{pmatrix} \quad (18.27)$$

The $(T^2)_{13} = T_{12}T_{23} = 1$ element of T^2 for the 3-disk transition matrix (17.9)

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}^2 = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}. \quad (18.28)$$

corresponds to path $3 \rightarrow 2 \rightarrow 1$, the only 2-step path from 3 to 1, while $(T^2)_{33} = T_{31}T_{13} + T_{32}T_{23} = 2$ counts the two returning, periodic paths $\overline{31}$ and $\overline{32}$.

To identify the 2-cycles you have to look at the trace of T^2 . The diagonal terms of T^2 are

$$\begin{pmatrix} T_{11}^2 + T_{12}T_{21} + T_{13}T_{31} & & \\ & T_{22}^2 + T_{21}T_{12} + T_{23}T_{32} & \\ & & T_{33}^2 + T_{31}T_{13} + T_{32}T_{23} \end{pmatrix} \\ = \begin{pmatrix} t_1^2 + t_{12} + t_{13} & & \\ & t_2^2 + t_{21} + t_{23} & \\ & & t_3^2 + t_{31} + t_{32} \end{pmatrix}.$$

For example, the ‘weight’ or ‘little trace’ associated with the 2-cycle $c = \overline{12}$ is $t_{12} = T_{12}T_{21}$. Note that $\text{tr } T^2 = (T^2)_{11} + (T^2)_{22} + (T^2)_{33} = t_1^2 + t_2^2 + t_3^2 + 2(t_{12} + t_{13} + t_{23})$ has a contribution from each 2-cycle $\overline{12}$, $\overline{13}$, $\overline{23}$ twice, one contribution for each periodic point. In general, if $T_{jj} \neq 0$, we also get the fixed point contributions $t_j^2 = T_{jj}^2$, in agreement with (18.6).

[return to p. 350](#)

Example 18.2. Traces for binary symbolic dynamics. For example, for the $[8 \times 8]$ transition matrix $T_{s_1 s_2 s_3, s_0 s_1 s_2}$ version of (17.11), or any refined partition $[2^n \times 2^n]$ transition matrix, n arbitrarily large, the periodic point $\overline{100}$ contributes $t_{100} = z^3 T_{\overline{100}, \overline{010}} T_{\overline{010}, \overline{001}} T_{\overline{001}, \overline{100}}$ to $z^3 \text{tr } T^3$. This product is manifestly cyclically invariant, $t_{100} = t_{010} = t_{001}$, so a prime cycle $p = \overline{001}$ of period 3 contributes 3 times, once for each periodic point along its orbit.

[exercise 14.7](#)

For the binary labeled non-wandering set the first few traces are given by (consult tables 18.1 and 18.2)

$$\begin{aligned} z \text{tr } T &= t_0 + t_1, \\ z^2 \text{tr } T^2 &= t_0^2 + t_1^2 + 2t_{10}, \\ z^3 \text{tr } T^3 &= t_0^3 + t_1^3 + 3t_{100} + 3t_{101}, \\ z^4 \text{tr } T^4 &= t_0^4 + t_1^4 + 2t_{10}^2 + 4t_{1000} + 4t_{1001} + 4t_{1011}. \end{aligned} \quad (18.29)$$

In the binary case the trace picks up only two contributions on the diagonal, $T_{0 \dots 0, 0 \dots 0} + T_{1 \dots 1, 1 \dots 1}$, no matter how much memory we assume. We can even take infinite memory

$M \rightarrow \infty$, in which case the contributing partitions are shrunk to the fixed points, $\text{tr } T = T_{0,0} + T_{1,1}$.

If there are no restrictions on symbols, the symbolic dynamics is complete, and *all* binary sequences are *admissible* (or *allowable*) itineraries. As this type of symbolic dynamics pops up frequently, we list the shortest binary prime cycles in table 18.1.

Example 18.3. Topological polynomial for complete binary dynamics. (Continuation of example 17.1) There are only two non-intersecting loops, yielding

exercise 14.2
return to p. 352



$$\det(1 - zT) = 1 - t_0 - t_1 - (t_{01} - t_0 t_1) = 1 - 2z \quad (18.30)$$

$$\begin{array}{c} \text{Diagram: A transition graph with two nodes, 0 and 1. Node 0 has a self-loop and an arrow to node 1. Node 1 has a self-loop and an arrow back to node 0.} \\ = 1 - \text{Diagram: Node 0 with a self-loop} - \text{Diagram: Node 1 with a self-loop} - \left(\text{Diagram: Path 0 to 1} - \text{Diagram: Path 1 to 0} \right) \end{array}$$

Due to the symmetry under $0 \leftrightarrow 1$ interchange, this is a redundant graph (the 2-cycle t_{01} is exactly shadowed by the 1-cycles). Another way to see is that itineraries are labeled by the $\{0, 1\}$ links, node labels can be omitted. As both nodes have 2 in-links and 2 out-links, they can be identified, and a more economical presentation is in terms of the $[1 \times 1]$ adjacency matrix (17.12)

$$\det(1 - zA) = 1 - t_0 - t_1 = 1 - 2z \quad (18.31)$$

$$\begin{array}{c} \text{Diagram: A single node with two self-loops.} \\ = 1 - \text{Diagram: Node 0 with a self-loop} - \text{Diagram: Node 1 with a self-loop} \end{array}$$

The leading (and only) zero of this characteristic polynomial yields the topological entropy $e^h = 2$. As there are $K_n = 2^n$ binary strings of length N , this comes as no surprise.

return to p. 356

Example 18.4. ‘Golden mean’ pruning. The ‘golden mean’ pruning of example 17.5 has one grammar rule: the substring `_11_` is forbidden. The corresponding transition graph non-intersecting loops are of length 1 and 2, so the topological polynomial is given by

exercise 18.5

$$\det(1 - zT) = 1 - t_0 - t_{01} = 1 - z - z^2 \quad (18.32)$$

$$\begin{array}{c} \text{Diagram: A transition graph with two nodes, 0 and 1. Node 0 has a self-loop and an arrow to node 1. Node 1 has a self-loop.} \\ = 1 - \text{Diagram: Node 0 with a self-loop} - \text{Diagram: Path 0 to 1} \end{array}$$

The leading root of this polynomial is the golden mean, so the entropy (18.4) is the logarithm of the golden mean, $h = \ln \frac{1+\sqrt{5}}{2}$.

return to p. 356

Example 18.5. Nontrivial pruning. The non-self-intersecting loops of the transition graph of figure 17.3 (d) are indicated in figure 17.3 (e). The determinant can be written down by inspection, as the sum of all possible partitions of the graph into products of non-intersecting loops, with each loop carrying a minus sign:

$$\begin{aligned} \det(1 - zT) = & 1 - t_0 - t_{0011} - t_{0001} - t_{00011} \\ & + t_0 t_{0011} + t_{0011} t_{0001} \end{aligned} \quad (18.33)$$

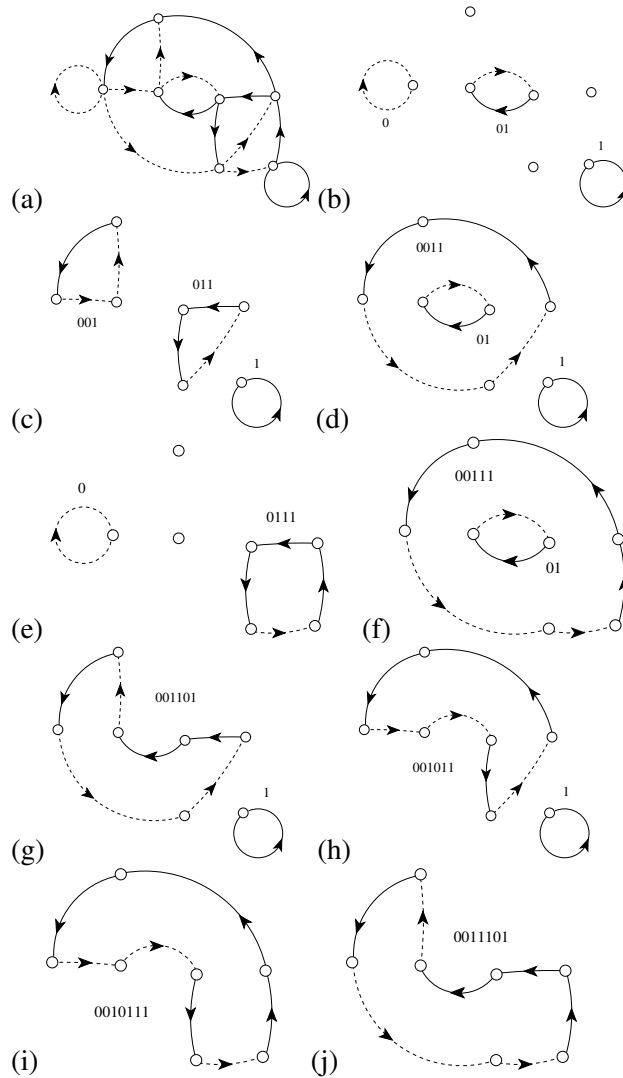


Figure 18.3: (a) The region labels in figure 17.4 have been omitted, since the links alone track the symbolic dynamics. (b)-(j) The fundamental cycles (18.36) for the transition graph (a), i.e., the set of its non-self-intersecting loops. Each loop represents a local trace t_p , as in (17.5). The isolated open dots are vertices not visited by a local trace.

With $t_p = z^{n_p}$, where n_p is the period of the p -cycle, the smallest root of

$$0 = 1 - z - 2z^4 + z^8 \quad (18.34)$$

yields the topological entropy $h = -\ln z$, $z = 0.658779 \dots$, $h = 0.417367 \dots$, significantly smaller than the entropy of the covering symbolic dynamics, the complete binary shift with topological entropy $h = \ln 2 = 0.693 \dots$

exercise 18.9
return to p. 356

Example 18.6. Loop expansion of a transition graph. (Continued from example 17.3) Consider a state space covered by 7 neighborhoods (17.8), with the topological time evolution given by the transition graph of figure 17.4.

The determinant $\det(1 - zT)$ of the transition graph in figure 17.4 can be read off the graph, and expanded as a polynomial in z , with coefficients given by products of non-

intersecting loops (traces of powers of T) of the transition graph figure 18.3:

$$\begin{aligned}
 \det(1 - zT) = & 1 - (t_0 + t_1)z - (t_{01} - t_0t_1)z^2 - (t_{001} + t_{011} - t_{01}t_0 - t_{01}t_1)z^3 \\
 & - (t_{0011} + t_{0111} - t_{001}t_1 - t_{011}t_0 - t_{011}t_1 + t_{01}t_0t_1)z^4 \\
 & - (t_{00111} - t_{0111}t_0 - t_{0011}t_1 + t_{011}t_0t_1)z^5 \\
 & - (t_{001011} + t_{001101} - t_{0011}t_0 - t_{001}t_{011})z^6 \\
 & - (t_{0010111} + t_{0011101} - t_{001011}t_1 - t_{001101}t_1 - t_{00111}t_{01} + t_{0011}t_{01}t_1 + t_{001}t_{011}t_1)z^7.
 \end{aligned} \tag{18.35}$$

Twelve cycles up to period 7 are fundamental cycles:

$$\bar{0}, \bar{1}, \bar{01}, \bar{001}, \bar{011}, \bar{0011}, \bar{0111}, \bar{00111}, \bar{001011}, \bar{001101}, \bar{0010111}, \bar{0011101}, \tag{18.36}$$

out of the total of 41 prime cycles (listed in table 18.1) up to cycle period 7. The topological polynomial $t_p \rightarrow z^{n_p}$

$$1/\zeta_{\text{top}}(z) = 1 - 2z - z^7$$

is interesting; the shadowing fails first at the cycle period $n = 7$, so the topological entropy is only a bit smaller than the binary $h = \ln 2$. Not exactly obvious from the partition (17.8).

[return to p. 356](#)

Example 18.7. Complete binary $\det(1 - zT)$ expansion. (Continuation of example 17.6) Consider the loop expansion of the binary 1-step memory transition graph (17.11)

[section 23.1.1](#)

$$\begin{aligned}
 \text{Diagram} &= 1 - \text{Diagram 1} - \text{Diagram 2} - \left(\text{Diagram 3} - \text{Diagram 4} - \text{Diagram 5} \right) \\
 &= 1 - t_0 - t_1 - [(t_{01} - t_1t_0)] - [(t_{001} - t_{01}t_0) + (t_{011} - t_{01}t_1)] \\
 &\quad - [(t_{0001} - t_0t_{001}) + (t_{0111} - t_{011}t_1) \\
 &\quad \quad + (t_{0011} - t_{001}t_1 - t_0t_{011} + t_0t_{01}t_1)] \\
 &= 1 - \sum_f t_f - \sum_n \hat{c}_n = 1 - 2z.
 \end{aligned} \tag{18.37}$$

(Continued in sect. 23.1.1.)

[return to p. 356](#)

Example 18.8. Complete N -ary dynamics. To check formula (18.24) for the finite-grammar situation, consider the complete N -ary dynamics (17.7) for which the number of periodic points of period n is simply $\text{tr } T_c^n = N^n$. Substituting

$$\sum_{n=1}^{\infty} \frac{z^n}{n} \text{tr } T_c^n = \sum_{n=1}^{\infty} \frac{(zN)^n}{n} = -\ln(1 - zN),$$

into (18.15) we verify (18.14). The logarithmic derivative formula (18.24) in this case does not buy us much either, it simply recovers

$$\sum_{n=1}^{\infty} N_n z^n = \frac{Nz}{1 - Nz}.$$

[return to p. 362](#)

Example 18.9. Nontrivial pruned dynamics. Consider the pruning of figure 17.3 (e). Substituting (18.24) we obtain

$$\sum_{n=1} N_n z^n = \frac{z + 8z^4 - 8z^8}{1 - z - 2z^4 + z^8}. \quad (18.38)$$

The topological zeta function is not merely a tool for extracting the asymptotic growth of N_n ; it also yields the exact numbers of periodic points. In case at hand it yields a 4-term recurrence $N_1 = N_2 = N_3 = 1$, $N_n = 2n + 1$ for $n = 4, 5, 6, 7$, $N_8 = 25$, and $N_n = N_{n-1} + 2N_{n-4} - N_{n-8}$ for $n > 8$.

[return to p. 362](#)

Example 18.10. Counting N -disk periodic points.



A simple example of pruning is the exclusion of “self-bounces” in the N -disk game of pinball. The number of points that are mapped back onto themselves after n iterations is given by $N_n = \text{tr } T^n$. The pruning of self-bounces eliminates the diagonal entries, $T_{N\text{-disk}} = T_c - \mathbf{1}$, so the number of the N -disk periodic points is

$$N_n = \text{tr } T_{N\text{-disk}}^n = (N - 1)^n + (-1)^n (N - 1). \quad (18.39)$$

Here T_c is the complete symbolic dynamics transition matrix (17.7). For the N -disk pruned case (18.39), Möbius inversion (18.26) yields

$$\begin{aligned} M_n^{N\text{-disk}} &= \frac{1}{n} \sum_{d|n} \mu\left(\frac{n}{d}\right) (N - 1)^d + \frac{N - 1}{n} \sum_{d|n} \mu\left(\frac{n}{d}\right) (-1)^d \\ &= M_n^{(N-1)} \quad \text{for } n > 2. \end{aligned} \quad (18.40)$$

There are no fixed points, so $M_1^{N\text{-disk}} = 0$. The number of periodic points of period 2 is $N^2 - N$, hence there are $M_2^{N\text{-disk}} = N(N - 1)/2$ prime cycles of period 2; for periods $n > 2$, the number of prime cycles is the same as for the complete $(N - 1)$ -ary dynamics of table 18.3.

[return to p. 363](#)

Example 18.11. Pruning individual cycles.



Consider the 3-disk game of pinball. The prohibition of repeating a symbol affects counting only for the fixed points and the 2-cycles. Everything else is the same as counting for a complete binary dynamics (18.40). To obtain the topological zeta function, just divide out the binary 1- and 2-cycles $(1 - zt_0)(1 - zt_1)(1 - z^2t_{01})$ and multiply with the correct 3-disk 2-cycles $(1 - z^2t_{12})(1 - z^2t_{13})(1 - z^2t_{23})$:

$$\begin{aligned} 1/\zeta_{3\text{-disk}} &= (1 - 2z) \frac{(1 - z^2)^3}{(1 - z)^2(1 - z^2)} \\ &= (1 - 2z)(1 + z)^2 = 1 - 3z^2 - 2z^3. \end{aligned} \quad (18.41)$$

[exercise 18.14](#)

[exercise 18.15](#)

The factorization reflects the underlying 3-disk symmetry; we shall rederive it in (25.32). As we shall see in chapter 25, symmetries lead to factorizations of topological polynomials and topological zeta functions.

[return to p. 363](#)

Example 18.12. Alphabet $\{a, cb^k; \bar{b}\}$. (Continuation of exercise 18.16) In the cycle counting case, the dynamics in terms of $a \rightarrow z$, $cb^k \rightarrow z + z^2 + z^3 + \cdots = z/(1 - z)$ is a complete binary dynamics with the explicit fixed point factor $(1 - t_b) = (1 - z)$:

[exercise 18.19](#)

$$1/\zeta_{\text{top}} = (1 - z) \left(1 - z - \frac{z}{1 - z} \right) = 1 - 3z + z^2.$$

return to p. 363

Example 18.13. Counting cat map periodic orbits. (Continued from example 14.12) The five sub-rectangles M_j of the two-rectangle Adler-Weiss partition of figure 14.18 (c) motivate introduction of a 5-letter alphabet

$$\bar{\mathcal{A}} = \{1, 2, 3, 4, 5\} = \{A^0A, B^1A, A^1A, B^0B, A^1B\}, \quad (18.42)$$

see figure 18.4 (b), which encodes the links of the transition graph of figure 14.18 (d). The loop expansion of the determinant (18.13) of the transition graph T of figure 18.4 (b) is given by all non-intersecting walks on the graph

$$\det(1 - zT) = 1 - z(t_1 + t_3 + t_4) - z^2(t_{25} - (t_1 + t_3)t_4), \quad (18.43)$$

where t_p are traces over fundamental cycles, the three fixed points $t_1 = T_{A^0A}$, $t_3 = T_{A^1A}$, $t_4 = T_{B^0B}$, and the 2-cycle $t_{25} = T_{B^1A}T_{A^1B}$.

As the simplest application, consider counting all admissible cat map periodic orbits. This is accomplished by setting the non-vanishing links of the transition graph to $T_{ji} = 1$, resulting in the $s = 3$ cat map topological zeta function

remark 14.5

$$1/\zeta_{\text{top}}(z) = \frac{1 - 3z + z^2}{(1 - z)^2}, \quad (18.44)$$

where the numerator $(1 - z)^2$ corrects the overcounting of the fixed point at the origin due to assigning it to both M_A (twice) and M_B rectangles.

The number of *periodic points* of period n is given by (18.24), the logarithmic derivative of the topological zeta function. Substituting the cat map topological zeta function (18.44) we obtain

$$\begin{aligned} \sum_{n=1} N_n z^n &= \frac{s - 2z}{1 - sz + z^2} - \frac{2}{1 - z} \\ &= z + 5z^2 + 16z^3 + 45z^4 + 121z^5 + 320z^6 + 841z^7 \\ &\quad + 2205z^8 + 5776z^9 + 15125z^{10} + 39601z^{11} + \dots \end{aligned} \quad (18.45)$$

The number of *prime cycles* is given by (18.26),

$$\begin{aligned} \sum_{n=1} M_n z^n &= z + 2z^2 + 5z^3 + 10z^4 + 24z^5 + 50z^6 + 120z^7 \\ &\quad + 270z^8 + 640z^9 + 1500z^{10} + 3600z^{11} + \dots, \end{aligned} \quad (18.46)$$

in agreement with the Bird and Vivaldi [4] census.

Above derivation was based on the explicit Adler-Weiss generating partition (18.43). However, no explicit generating partition is needed to count cat map periodic orbits. Cat map (8.52) periodic points are the fixed points of

$$\begin{bmatrix} q_t \\ p_t \end{bmatrix} = \begin{bmatrix} q_{t+n} \\ p_{t+n} \end{bmatrix} = J^n \begin{bmatrix} q_t \\ p_t \end{bmatrix} \pmod{1},$$

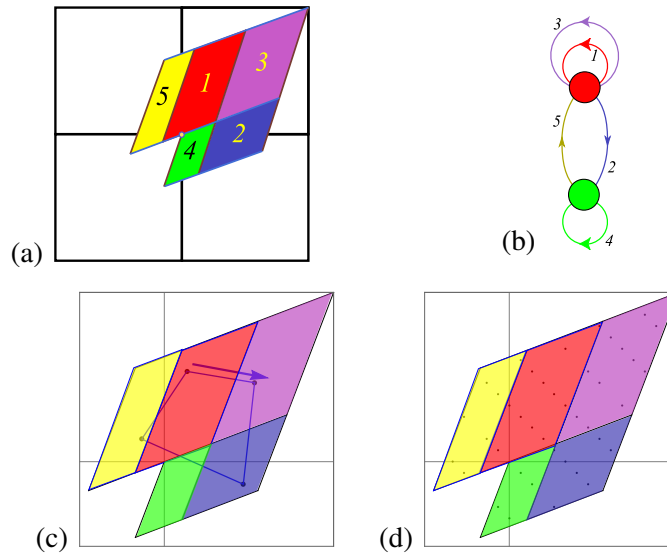


Figure 18.4: (Continued from figure 14.18) (a) The sub-rectangles M_i of figure 14.18(c). (b) Admissible orbits correspond to walks on the transition graph of figure 14.18(d), with rectangles M_A (red) and M_B (green) as nodes, and the links labeled by 5-letter alphabet (18.42), see the loop expansion (18.43). (c) The $\overline{0111}$ is an example of a period-4 prime cycle. (d) All prime period-4 periodic points, see the periodic points, one-per prime orbit (18.50). Note that they all land in the partition (a).

so on the unwrapped phase space lattice, tiled by repeats of the unit square of the cat map torus,

$$(J^n - \mathbf{1}) \begin{bmatrix} q_t \\ p_t \end{bmatrix} = \begin{bmatrix} m_t^q \\ m_t^p \end{bmatrix}, \quad (m_t^q, m_t^p) \in \mathbb{Z}^2, \quad (18.47)$$

matrix $(J^n - \mathbf{1})$ stretches the unit square into the ‘fundamental parallelepiped’ whose volume equals the number of periodic points (continued in example 18.14).

(H. Liang and P. Cvitanović)

[return to p. 357](#)

Example 18.14. Cat map period-4 prime cycles. (Continued from example 18.13) As a hands-on example, let us count the $M_4 = 10$ admissible prime 4-cycles. The admissible blocks M_p can be read off as walks on either the 5-letter alphabet (18.42) graph, see figure 18.4(b), or the 3-letter alphabet (14.26) graph, see figure 14.18(d). They are, in 5-letter (top), and 3-letter (bottom) alphabets

$$\begin{array}{ccccc} \overline{1113} & \overline{1125} & \overline{1245} & \overline{1253} & \overline{1325} \\ 0001 & 0011 & 0101 & 0111 & 0111 \\ \hline \overline{1133} & \overline{3325} & \overline{3331} & \overline{3245} & \overline{4452} \\ 0011 & 1111 & 1110 & 1101 & 0011 \end{array}. \quad (18.48)$$

As the relation between the symbol blocks M and the corresponding orbit Φ_M is linear, for M an admissible symbol block, the corresponding orbit Φ_M is given by the Green’s function

$$\Phi_M = dM, \quad g = \frac{1}{-d + s\mathbf{1} - d^{-1}}, \quad (18.49)$$

Thus the corresponding periodic orbits Φ_p are computed using Green's function (18.49) (the inverse of the $[4 \times 4]$ orbit Jacobian matrix, easiest to evaluate by discrete Fourier transforms, see sect. 18.43):

$$M_{0001} \Rightarrow \Phi_{0001} = g \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{15} \begin{bmatrix} 3 \\ 2 \\ 3 \\ 7 \end{bmatrix}.$$

Likewise,

$$\begin{aligned} \Phi_{001\bar{1}}^\top &= \frac{1}{15} \begin{bmatrix} -1 & 1 & 4 & -4 \end{bmatrix}, & \Phi_{010\bar{1}}^\top &= \frac{1}{15} \begin{bmatrix} 0 & 5 & 0 & -5 \end{bmatrix} \\ \Phi_{011\bar{1}}^\top &= \frac{1}{15} \begin{bmatrix} 4 & 6 & -1 & 6 \end{bmatrix}, & \Phi_{011\bar{1}}^\top &= \frac{1}{15} \begin{bmatrix} 2 & 8 & 7 & -2 \end{bmatrix} \\ \Phi_{0011}^\top &= \frac{1}{15} \begin{bmatrix} 5 & 5 & 10 & 10 \end{bmatrix}, & \Phi_{111\bar{1}}^\top &= \frac{1}{15} \begin{bmatrix} 9 & 11 & 9 & 1 \end{bmatrix} \\ \Phi_{1110}^\top &= \frac{1}{15} \begin{bmatrix} 12 & 13 & 12 & 8 \end{bmatrix}, & \Phi_{110\bar{1}}^\top &= \frac{1}{15} \begin{bmatrix} 7 & 8 & 2 & -2 \end{bmatrix} \\ \Phi_{00\bar{1}1}^\top &= \frac{1}{15} \begin{bmatrix} 1 & -1 & -4 & 4 \end{bmatrix}. \end{aligned} \quad (18.50)$$

As demonstrated in figure 18.4(d), each of these 10 prime 4-cycles' periodic points (x_t, x_{t+1}) visit the rectangles $\mathcal{M}_A$ or $\mathcal{M}_B$ of figure 14.18(b) in the temporal order dictated by the transition graph, and are thus all admissible cycles. (H. Liang and P. Cvitanović)

[return to p. 357](#)

Exercises

18.1. A transition matrix for 3-disk pinball.

- Draw the transition graph corresponding to the 3-disk ternary symbolic dynamics, and write down the corresponding transition matrix corresponding to the graph. Show that iteration of the transition matrix results in two coupled linear difference equations, - one for the diagonal and one for the off diagonal elements. (Hint: relate $\text{tr } T^n$ to $\text{tr } T^{n-1} + \dots$)
- Solve the above difference equation and obtain the number of periodic orbits of period n . Compare your result with table 18.4.
- Find the eigenvalues of the transition matrix T for the 3-disk system with ternary symbolic dynamics and calculate the topological entropy. Compare this to the topological entropy obtained from the binary symbolic dynamics $\{0, 1\}$.

18.2. 3-disk prime cycle counting.

A *prime cycle* p of period n_p is a single traversal of the orbit; its label is a non-repeating symbol string of n_p symbols. For example, $\overline{12}$ is prime, but $\overline{2121}$ is not, since it is $\overline{21} = \overline{12}$ repeated.

Verify that a 3-disk pinball has 3, 2, 3, 6, 9, $\dots$ prime cycles of period 2, 3, 4, 5, 6, $\dots$.

18.3. Sum of A_{ij} is like a trace.

Let A be a matrix with eigenvalues λ_k . Show that

$$\Gamma_n := \sum_{i,j} [A^n]_{ij} = \sum_k c_k \lambda_k^n.$$

- Under what conditions do $\ln |\text{tr } A^n|$ and $\ln |\Gamma_n|$ have the same asymptotic behavior as $n \rightarrow \infty$, i.e., their ratio converges to one?
- Do eigenvalues λ_k need to be distinct, $\lambda_k \neq \lambda_l$ for $k \neq l$? How would a degeneracy $\lambda_k = \lambda_l$ affect your argument for (a)?

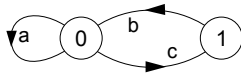
18.4. Loop expansions.

Prove by induction the sign rule in the determinant expansion (18.13):

$$\det(1 - zT) = \sum_{k \geq 0} \sum_{p_1 + \dots + p_k} (-1)^k t_{p_1} t_{p_2} \dots t_{p_k}.$$

18.5. Transition matrix and cycle counting.

Suppose you are given the transition graph



This diagram can be encoded by a matrix T , where the entry T_{ij} means that there is a link connecting node i to node j . The value of the entry is the weight of the link.

- Walks on the graph are given a weight that is the product of the weights of all links crossed by the walk. Convince yourself that the transition matrix for this graph is:

$$T = \begin{bmatrix} a & b \\ c & 0 \end{bmatrix}.$$

- Enumerate all the walks of length three on the transition graph. Now compute T^3 and look at the entries. Is there any relation between the terms in T^3 and all the walks?

- (c) Show that T_{ij}^n is the number of walks from point i to point j in n steps. (Hint: one might use the method of induction.)
- (d) Estimate the number K_n of walks of length n for this simple transition graph.
- (e) The topological entropy h measures the rate of exponential growth of the total number of walks K_n as a function of n . What is the topological entropy for this transition graph?

18.6. **Alphabet {0,1}, prune _00_.** The transition graph example 17.8 implements this pruning rule which implies that “0” must always be bracketed by “1”s; in terms of a new symbol $2 := 10$, the dynamics becomes unrestricted symbolic dynamics with binary alphabet $\{1,2\}$. The cycle expansion (18.13) becomes

$$\begin{aligned} 1/\zeta &= (1 - t_1)(1 - t_2)(1 - t_{12})(1 - t_{112}) \dots \\ &= 1 - t_1 - t_2 - (t_{12} - t_1 t_2) \\ &\quad - (t_{112} - t_{12} t_1) - (t_{122} - t_{12} t_2) \dots \end{aligned} \quad (18.51)$$

In the original binary alphabet this corresponds to:

$$\begin{aligned} 1/\zeta &= 1 - t_1 - t_{10} - (t_{110} - t_1 t_{10}) \\ &\quad - (t_{1110} - t_{110} t_1) - (t_{11010} - t_{110} t_{10}) \dots \end{aligned} \quad (18.52)$$

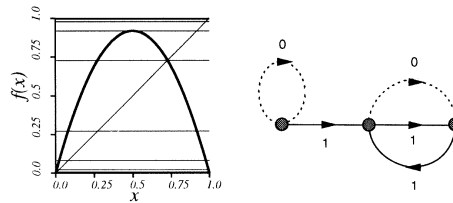
This symbolic dynamics describes, for example, circle maps with the golden mean winding number. For unimodal maps this symbolic dynamics is realized by the tent map of exercise 14.6.

18.7. **‘Golden mean’ pruned map.** (continuation of exercise 14.6) Show that the total number of periodic orbits of period n for the ‘golden mean’ tent map is

$$\frac{(1 + \sqrt{5})^n + (1 - \sqrt{5})^n}{2^n}.$$

Continued in exercise 22.1. See also exercise 18.8.

18.8. **A unimodal map with ‘golden mean’ pruning.** Consider the unimodal map



for which the critical point maps into the right hand fixed point in three iterations, $S^+ = 100\bar{1}$. Show that the admissible itineraries are generated by the above transition graph, with transient neighborhood of $\bar{0}$ fixed point, and _00_ pruned from the recurrent set. (K.T. Hansen)

18.9. **Glitches in shadowing.** (medium difficulty) Note that the combination t_{00011} minus the “shadow” $t_0 t_{0011}$ in (18.33) cancels exactly, and does not contribute to the topological zeta function (18.34). Are you able to construct a smaller transition graph than figure 17.3 (e)?

18.10. **Whence Möbius function?** To understand the origin of the Möbius function (18.26), consider the function

$$f(n) = \sum_{d|n} g(d) \quad (18.53)$$

where $d|n$ stands for sum over all divisors d of n . Invert recursively this infinite tower of equations and derive the *Möbius inversion formula*

$$g(n) = \sum_{d|n} \mu(n/d) f(d). \quad (18.54)$$

- 18.11. **Counting prime binary cycles.** In order to get comfortable with Möbius inversion reproduce the results of the second column of table 18.3.

Write a program that determines the number of prime cycles of period n . You might want to have this program later on to be sure that you have missed no 3-pinball prime cycles.

- 18.12. **Counting subsets of cycles.** The techniques developed above can be generalized to counting subsets of cycles. Consider the simplest example of a dynamical system with a complete binary tree, a repeller map (14.21) with two straight branches, which we label 0 and 1. Every cycle weight for such map factorizes, with a factor t_0 for each 0, and factor t_1 for each 1 in its symbol string. Prove that the transition matrix traces (18.29) collapse to $\text{tr}(T^k) = (t_0 + t_1)^k$, and $1/\zeta$ is simply

$$\prod_p (1 - t_p) = 1 - t_0 - t_1 \quad (18.55)$$

Substituting (18.55) into the identity

$$\prod_p (1 + t_p) = \prod_p \frac{1 - t_p^2}{1 - t_p}$$

we obtain

$$\begin{aligned} \prod_p (1 + t_p) &= \frac{1 - t_0^2 - t_1^2}{1 - t_0 - t_1} \\ &= 1 + t_0 + t_1 + \frac{2t_0t_1}{1 - t_0 - t_1} \\ &= 1 + t_0 + t_1 \\ &\quad + \sum_{n=2}^{\infty} \sum_{k=1}^{n-1} 2 \binom{n-2}{k-1} t_0^k t_1^{n-k}. \end{aligned}$$

Hence for $n \geq 2$ the number of terms in the cumulant expansion with k 0's and $n - k$ 1's in their symbol sequences is $2 \binom{n-2}{k-1}$.

In order to count the number of prime cycles in each such subset we denote with $M_{n,k}$ ($n = 1, 2, \dots$; $k = \{0, 1\}$ for $n = 1$; $k = 1, \dots, n - 1$ for $n \geq 2$) the number of prime n -cycles whose labels contain k zeros. Show that

$$\begin{aligned} M_{1,0} &= M_{1,1} = 1, \quad n \geq 2, k = 1, \dots, n - 1 \\ nM_{n,k} &= \sum_{m \mid \frac{n}{k}} \mu(m) \binom{n/m}{k/m} \end{aligned}$$

where the sum is over all m which divide both n and k . (continued as exercise 23.7)

- 18.13. **Logarithmic periodicity of $\ln N_n$.** (medium difficulty) Plot $(\ln N_n, nh)$ for a system with a nontrivial finite transition graph. Do you see any periodicity? If yes, why?
- 18.14. **Symmetric 4-disk pinball topological zeta function.** Show that the 4-disk pinball topological zeta function (the pruning affects only the fixed points and the 2-cycles) is given

by

$$\begin{aligned}
 1/\zeta_{\text{top}}^{4\text{-disk}} &= (1-3z) \frac{(1-z^2)^6}{(1-z)^3(1-z^2)^3} \\
 &= (1-3z)(1+z)^3 \\
 &= 1-6z^2-8z^3-3z^4.
 \end{aligned} \tag{18.56}$$

- 18.15. **Symmetric N -disk pinball topological zeta function.** Show that for an N -disk pinball, the topological zeta function is given by

$$\begin{aligned}
 1/\zeta_{\text{top}}^{N\text{-disk}} &= (1-(N-1)z) \times \\
 &\quad \frac{(1-z^2)^{N(N-1)/2}}{(1-z)^{N-1}(1-z^2)^{(N-1)(N-2)/2}} \\
 &= (1-(N-1)z)(1+z)^{N-1}.
 \end{aligned} \tag{18.57}$$

The topological zeta function has a root $z^{-1} = N-1$, as we already know it should from (18.39) or (18.14). We shall see in sect. 25.4 that the other roots reflect the symmetry factorizations of zeta functions.

- 18.16. **Alphabet $\{a, b, c\}$, prune $_{ab}$.** Write down the topological zeta function for this pruning rule.
- 18.17. **Alphabet $\{0,1\}$, prune n repeats of “0” $_{000\dots 00}$.** This is equivalent to the n symbol alphabet $\{1, 2, \dots, n\}$ unrestricted symbolic dynamics, with symbols corresponding to the possible $10\dots 00$ block lengths: $2:=10$, $3:=100$, $\dots$, $n:=100\dots 00$. Show that the cycle expansion (18.13) becomes

$$\begin{aligned}
 1/\zeta &= 1 - t_1 - t_2 \cdots - t_n - (t_{12} - t_1 t_2) \cdots \\
 &\quad - (t_{1n} - t_1 t_n) \cdots
 \end{aligned}$$

- 18.18. **Alphabet $\{0,1\}$, prune $_{1000}, _{00100}, _{01100}$.** Show that the topological zeta function is given by

$$1/\zeta = (1-t_0)(1-t_1-t_2-t_{23}-t_{113}) \tag{18.58}$$

with the unrestricted 4-letter alphabet $\{1, 2, \underline{23}, \underline{113}\}$. Here 2 and 3 refer to 10 and 100 respectively, as in exercise 18.17.

- 18.19. **Alphabet $\{0,1\}$, prune $_{1000}, _{00100}, _{01100}, _{10011}$.** (This grammar arises from Hénon map pruning, see remark 15.3.) The first three pruning rules were incorporated in the preceeding exercise.

(a) Show that the last pruning rule $_{10011}$ leads (in a way similar to exercise 18.18) to the alphabet $\{\underline{21^k}, \underline{23}, \underline{21^k 113}; \bar{1}, \bar{0}\}$, and the cycle expansion

$$1/\zeta = (1-t_0)(1-t_1-t_2-t_{23}+t_1 t_{23}-t_{2113}). \tag{18.59}$$

Note that this says that 1, 23, 2, 2113 are the fundamental cycles; not all cycles up to period 7 are needed, only 2113.

(b) Show that the topological zeta function is

$$1/\zeta_{\text{top}} = (1-z)(1-z-z^2-z^5+z^6-z^7) \tag{18.60}$$

and that it yields the entropy $h = 0.522737642\dots$

- 18.20. **Alphabet {0,1}, prune only the fixed point $\bar{0}$.** This is equivalent to the *infinite* alphabet $\{1, 2, 3, 4, \dots\}$ unrestricted symbolic dynamics. The prime cycles are labeled by all non-repeating sequences of integers, ordered lexically: $t_n, n > 0; t_{mn}, t_{mmn}, \dots, n > m > 0; t_{mnr}, r > n > m > 0, \dots$ (see sect. 29.3). Now the number of fundamental cycles is infinite as well:

$$\begin{aligned}
 1/\zeta &= 1 - \sum_{n>0} t_n - \sum_{n>m>0} (t_{mn} - t_n t_m) \\
 &\quad - \sum_{n>m>0} (t_{mmn} - t_m t_{mn}) \\
 &\quad - \sum_{n>m>0} (t_{mnn} - t_{mn} t_n) \\
 &\quad - \sum_{r>n>m>0} (t_{mnr} + t_{mrn} - t_{mn} t_r \\
 &\quad \quad - t_{mr} t_n - t_m t_{nr} + t_m t_n t_r) \cdots
 \end{aligned} \tag{18.61}$$

. As shown in table 29.1, this grammar plays an important role in description of fixed points of marginal stability.

- 18.21. **Fundamental cycles.**



In sect. 18.3 we referred to the set of all non-self-intersecting loops $\{t_{p_1}, t_{p_2}, \dots, t_{p_f}\}$ as the *fundamental cycles*. This is not a very good definition, as transition graphs are not unique. If you have a better way to define the “fundamental cycles,” yours, or something discussed in the literature, let us know.

Chapter 19

Transporting densities

Paulina: I'll draw the curtain:
My lord's almost so far transported that
He'll think anon it lives.
—W. Shakespeare, *The Winter's Tale*

(P. Cvitanović, R. Artuso, L. Rondoni, and E.A. Spiegel)

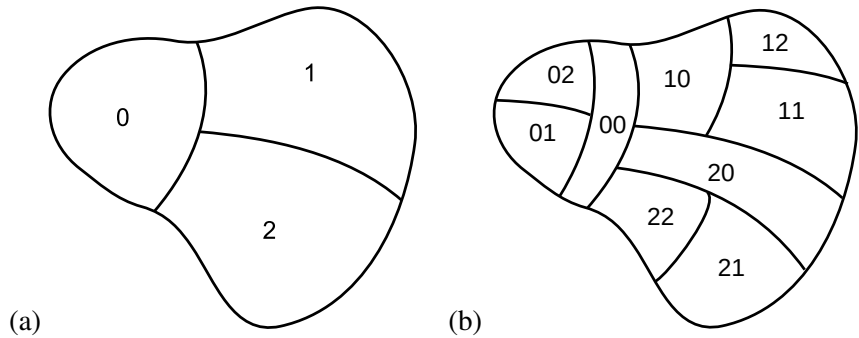
IN CHAPTERS 2, 3, 8 and 9 we learned how to track an individual trajectory, and saw that such a trajectory can be very complicated. In chapter 4 we studied a small neighborhood of a trajectory and learned that such neighborhood can grow exponentially with time, making the concept of tracking an individual trajectory for long times a purely mathematical idealization.

While the trajectory of an individual representative point may be highly convoluted, as we shall see, the density of these points might evolve in a manner that is relatively smooth. The evolution of the density of representative points is for this reason (and other that will emerge in due course) of great interest. So are the behaviors of other properties carried by the evolving swarm of representative points.

We shall now show that the global evolution of the density of representative points is conveniently formulated in terms of linear action of evolution operators. We shall also show that the important, long-time “natural” invariant densities are unspeakably unfriendly and essentially uncomputable everywhere singular functions with support on fractal sets. Hence, in chapter 20 we rethink what is it that the theory needs to predict (“expectation values” of “observables”), relate these to the eigenvalues of evolution operators, and in chapters 21 to 23 show how to compute these without ever having to compute a “natural” invariant density ρ_0 .



Figure 19.1: (a) First level of partitioning: A coarse partition of $\mathcal{M}$ into regions $\mathcal{M}_0$, $\mathcal{M}_1$, and $\mathcal{M}_2$. (b) $n = 2$ level of partitioning: A refinement of the above partition, with each region $\mathcal{M}_i$ subdivided into $\mathcal{M}_{i0}$, $\mathcal{M}_{i1}$, and $\mathcal{M}_{i2}$.



19.1 Measures

Do I then measure, O my God, and know not what I measure?



—St. Augustine, *The confessions of Saint Augustine*

A fundamental concept in the description of dynamics of a chaotic system is that of *measure*, which we denote by $d\mu(x) = \rho(x) dx$. An intuitive way to define and construct a physically meaningful measure is by a process of *coarse-graining*. Consider a sequence $1, 2, \dots, n, \dots$ of increasingly refined partitions of state space, figure 19.1, into 3 regions $\mathcal{M}_i$ defined by the indicator (characteristic function) $[\mathcal{A}]$ (or $\mathbb{1}_{\mathcal{A}}$)

$$[\mathcal{A}](x) = \mathbb{1}_{\mathcal{A}}(x) = \begin{cases} 1 & \text{if } x \in \mathcal{A} \\ 0 & \text{otherwise} \end{cases} . \quad (19.1)$$

A coarse-grained measure is obtained by assigning the “mass,” or the fraction of trajectories contained in the i th region $\mathcal{M}_i \subset \mathcal{M}$ at the n th level of partitioning of the state space:

$$\Delta\mu_i = \int_{\mathcal{M}} d\mu(x) [\mathcal{M}_i](x) = \int_{\mathcal{M}_i} d\mu(x) = \int_{\mathcal{M}_i} dx \rho(x) . \quad (19.2)$$

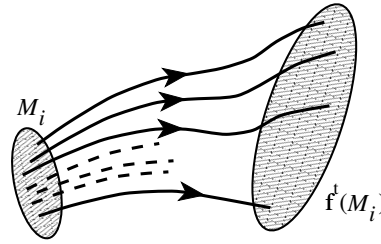
The function $\rho(x) = \rho(x, t)$ denotes the *density* of representative points in state space at time t . This density can be (and in chaotic dynamics, often is) an arbitrarily ugly function, and it may display remarkable singularities; for instance, there may exist directions along which the measure is singular with respect to the Lebesgue measure. We shall assume that the measure is normalized

$$\sum_i^{(n)} \Delta\mu_i = 1 , \quad (19.3)$$

where the sum is over subregions i at the n th level of partitioning. The infinitesimal measure $\rho(x) dx$ can be thought of as an infinitely refined partition limit of $\Delta\mu_i = |\mathcal{M}_i| \rho(x_i)$ (see (19.11)), where $|\mathcal{M}_i|$ is the volume of subregion $\mathcal{M}_i$, $x_i \in \mathcal{M}_i$ is any point in the subregion, and all $|\mathcal{M}_i| \rightarrow 0$ as $n \rightarrow \infty$. It follows from (19.3) that $\rho(x)$ is normalized probability density

$$\int_{\mathcal{M}} dx \rho(x) = 1 . \quad (19.4)$$

Figure 19.2: The evolution rule f^t can be used to map a region M_i of the state space into the region $f^t(M_i)$.



So far, any arbitrary sequence of partitions will do. What are intelligent ways of partitioning state space? We already know the answer from chapter 14, but let us anyway have another look at this, in order to develop some intuition about how the dynamics transports densities.

chapter 14

19.2 Perron-Frobenius operator

Given a density, the question arises as to what it might evolve into with time. Consider a swarm of representative points making up the measure contained in a region M_i at time $t = 0$. As the flow evolves, this region is carried into $f^t(M_i)$, as in figure 19.2. No trajectory is created or destroyed, so the conservation of representative points requires that

$$\int_{f^t(M_i)} dx \rho(x, t) = \int_{M_i} dx_0 \rho(x_0, 0).$$

Transform the integration variable in the expression on the left hand side to the initial points $x_0 = f^{-t}(x)$,

$$\int_{M_i} dx_0 \rho(f^t(x_0), t) |\det J^t(x_0)| = \int_{M_i} dx_0 \rho(x_0, 0).$$

The density changes with time as the inverse of the Jacobian (4.28)

$$\rho(x, t) = \frac{\rho(x_0, 0)}{|\det J^t(x_0)|}, \quad x = f^t(x_0), \quad (19.5)$$

which makes sense: the density varies inversely with the infinitesimal volume occupied by the trajectories of the flow.



The relation (19.5) is linear in ρ , so the manner in which a flow transports densities may be recast into the language of operators, by writing

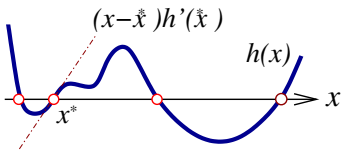
exercise 19.1

$$\rho(x, t) = (\mathcal{L}^t \circ \rho)(x) = \int_{\mathcal{M}} dx_0 \delta(x - f^t(x_0)) \rho(x_0, 0). \quad (19.6)$$

Let us check this formula. As long as the zero is not smack on the border of $\partial\mathcal{M}$, integrating Dirac delta functions is easy: $\int_{\mathcal{M}} dx \delta(x) = 1$ if $0 \in \mathcal{M}$, zero otherwise. The integral over a 1-dimensional Dirac delta function picks up the Jacobian of its



argument evaluated at all of its zeros:

$$\begin{aligned} \int dx \delta(h(x)) &= \int dx \delta(h(x)) = \sum_{\{x: h(x)=0\}} \frac{1}{|h'(x)|}, \end{aligned} \quad (19.7)$$


and in d dimensions the denominator is replaced by

$$\int dx \delta(h(x)) = \sum_j \int_{M_j} dx \delta(h(x)) = \sum_j \frac{1}{\left| \det \frac{\partial h(x_j)}{\partial x} \right|}, \quad (19.8)$$

where M_j is any open neighborhood that contains the single x_j zero of h . Now you can check that (19.6) is just a rewrite of (19.5):

exercise 19.2

$$\begin{aligned} (\mathcal{L}^t \circ \rho)(x) &= \sum_{x_0=f^{-t}(x)} \frac{\rho(x_0)}{|f^t(x_0)'|} \quad (1\text{-dimensional}) \\ &= \sum_{x_0=f^{-t}(x)} \frac{\rho(x_0)}{|\det J^t(x_0)|} \quad (d\text{-dimensional}). \end{aligned} \quad (19.9)$$

For a deterministic, invertible flow x has only one preimage x_0 ; allowing for multiple preimages also takes account of noninvertible mappings such as the ‘stretch & fold’ maps of the interval, discussed briefly in example 19.1, and in more detail in sect. 14.3.

We shall refer to the integral operator with the Dirac delta kernel (19.6) as the *Perron-Frobenius operator*:

$$\mathcal{L}^t(y, x) = \delta(y - f^t(x)). \quad (19.10)$$

exercise 19.3

example 28.7

The Perron-Frobenius operator assembles the density $\rho(y, t)$ at time t by going back in time to the density $\rho(x, 0)$ at time $t = 0$. The family of Perron-Frobenius operators $\{\mathcal{L}^t\}_{t \in \mathbb{R}_+}$ forms a semigroup parameterized by time

- (a) $\mathcal{L}^0 = I$
- (b) $\mathcal{L}^t \mathcal{L}^{t'} = \mathcal{L}^{t+t'} \quad t, t' \geq 0 \quad (\text{semigroup property})$.

If you do not like the word “kernel” you might prefer to think of $\mathcal{L}^t(y, x)$ as a matrix with indices x, y , and index summation in matrix multiplication replaced by an integral over x , $(\mathcal{L}^t \circ \rho)(y) = \int dx \mathcal{L}^t(y, x) \rho(x)$. In example 19.1, Perron-Frobenius operator is a matrix, and (19.12) illustrates a matrix approximation to the Perron-Frobenius operator.

remark 22.4



example 19.1
p. 398



fast track:
sect. 19.4, p. 387

19.3 Why not just leave it to a computer?

Another subtlety in the [dynamical systems] theory is that topological and measure-theoretic concepts of genericity lead to different results.

— John Guckenheimer

(R. Artuso and P. Cvitanović)

To a student with a practical bent the above Example 19.1 suggests a strategy for constructing evolution operators for smooth maps, as limits of partitions of state space into regions $\mathcal{M}_i$, with a piecewise-linear approximations f_i to the dynamics in each region, but that would be too naive; much of the physically interesting spectrum would be missed. As we shall see, the choice of function space for ρ is crucial, and the physically motivated choice is a space of smooth functions, rather than the space of piecewise constant functions.



chapter 28

All of the insight gained in this chapter and in what is to follow is nothing but an elegant way of thinking of the evolution operator, $\mathcal{L}$, as a matrix (this point of view will be further elaborated in chapter 28). There are many textbook methods of approximating an operator $\mathcal{L}$ by sequences of finite matrix approximations $\mathcal{L}$, but in what follows the great achievement will be that we shall avoid constructing any matrix approximation to $\mathcal{L}$ altogether. Why a new method? Why not just run it on a computer, as many do with such relish in diagonalizing quantum Hamiltonians?

The simplest possible way of introducing a state space discretization, figure 19.3, is to partition the state space $\mathcal{M}$ into a non-overlapping collection of sets $\mathcal{M}_j$, $j = 1, \dots, N$, and to consider densities (19.2) piecewise constant on each $\mathcal{M}_j$:

$$\rho(x) = \sum_{j=1}^N \rho_j \frac{[\mathcal{M}_j](x)}{|\mathcal{M}_j|} \quad (19.11)$$

where $[\mathcal{M}_j](x)$ is the indicator (19.1) of the set $\mathcal{M}_j$.

An example of a dynamically motivated piecewise constant measure, particularly easy to implement numerically, is the *equipartition* or *cylinder* measure. While an admissible infinite itinerary corresponds to a unique point in the state space, any finite itinerary $b = s_{t-\ell+1} \cdots s_{-1} s_0 s_1 s_2 \cdots s_t$ determines a *cylinder set* $\mathcal{M}_b$, the set of all points in $\mathcal{M}$ whose itineraries (14.14) share the same finite b symbol block, and arbitrary $a_i \in \mathcal{A}$. The cylinder measure ρ_b weighs all symbol sequences b of the same length n equally.

Such piecewise-constant densities are coarse grained presentations of fine grained density $\hat{\rho}(x)$, with (19.2)

$$\rho_i = \int_{\mathcal{M}_i} dx \hat{\rho}(x).$$

BRUTO INSENSITIVO METHOD:

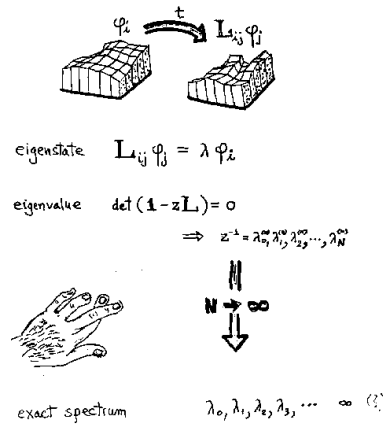


Figure 19.3: State space discretization approach to computing averages.

The Perron-Frobenius operator does not preserve the piecewise constant form, but we may reapply coarse graining to the evolved measure

$$\begin{aligned} \rho'_i &= \int_{\mathcal{M}_i} dx (\mathcal{L} \circ \rho)(x) \\ &= \sum_{j=1}^N \frac{\rho_j}{|\mathcal{M}_j|} \int_{\mathcal{M}_i} dx \int_{\mathcal{M}_j} dy \delta(x - f(y)) , \end{aligned}$$

or

$$\rho'_i = \sum_{j=1}^N \rho_j \frac{|\mathcal{M}_i \cap f^{-1}(\mathcal{M}_j)|}{|\mathcal{M}_j|} .$$

In this way

$$L_{ij} = \frac{|\mathcal{M}_i \cap f^{-1}(\mathcal{M}_j)|}{|\mathcal{M}_i|} , \quad \rho' = \rho L \quad (19.12)$$

is a matrix approximation to the Perron-Frobenius operator, and its leading left eigenvector is a piecewise constant approximation to the invariant measure.

remark 19.3

The problem with such state space discretization approaches is that they are blind, the grid knows not what parts of the state space are more or less important. This observation motivated the development of the invariant partitions of chaotic systems undertaken in chapter 14, we exploited the intrinsic topology of a flow to give us both an invariant partition of the state space and a measure of the partition volumes, in the spirit of figure 15.13.

Furthermore, a piecewise constant ρ belongs to an unphysical function space, and with such approximations one is plagued by numerical artifacts such as spurious eigenvalues. In chapter 28 we shall employ a more refined approach to extracting spectra, by expanding the initial and final densities ρ, ρ' in some basis $\varphi_0, \varphi_1, \varphi_2, \dots$ (orthogonal polynomials, let us say), and replacing $\mathcal{L}(y, x)$ by its

φ_α basis representation $\mathbf{L}_{\alpha\beta} = \langle \varphi_\alpha | \mathcal{L} | \varphi_\beta \rangle$. The art is then the subtle art of finding a “good” basis for which finite truncations of $\mathbf{L}_{\alpha\beta}$ give accurate estimates of the eigenvalues of $\mathcal{L}$.

chapter 28

Regardless of how sophisticated the choice of basis might be, the basic problem cannot be avoided - as illustrated by the natural measure for the Hénon map (3.18) sketched in figure 19.4, eigenfunctions of $\mathcal{L}$ are complicated, singular functions concentrated on fractal sets, and in general cannot be represented by a nice basis set of smooth functions. We shall resort to matrix representations of $\mathcal{L}$ and the φ_α basis approach only insofar this helps us prove that the spectrum that we compute is indeed the correct one, and that finite periodic orbit truncations do converge.



in depth:
chapter 1, p. 2

19.4 Invariant measures



A *stationary* or *invariant density* is a density left unchanged by the flow

$$\rho(x, t) = \rho(x, 0) = \rho(x). \quad (19.13)$$

Conversely, if such a density exists, the transformation $f^t(x)$ is said to be *measure-preserving*. As we are given deterministic dynamics and our goal is the computation of asymptotic averages of observables, our task is to identify interesting invariant measures for a given $f^t(x)$. Invariant measures remain unaffected by dynamics, so they are fixed points (in the infinite-dimensional function space of ρ densities) of the Perron-Frobenius operator (19.10), with the unit eigenvalue:

exercise 19.3

$$\mathcal{L}^t \rho(x) = \int_{\mathcal{M}} dy \delta(x - f^t(y)) \rho(y) = \rho(x). \quad (19.14)$$

We will construct explicitly such eigenfunction for the piecewise linear map in example 20.4, with $\rho(y) = \text{const}$ and eigenvalue 1. In general, depending on the choice of $f^t(x)$ and the function space for $\rho(x)$, there may be no, one, or many solutions of the eigenfunction condition (19.14). For instance, a singular measure $d\mu(x) = \delta(x - x_q)dx$ concentrated on an equilibrium point $x_q = f^t(x_q)$, or any linear combination of such measures, each concentrated on a different equilibrium point, is stationary. There are thus infinitely many stationary measures that can be constructed. Almost all of them are unnatural in the sense that the slightest perturbation will destroy them.

From a physical point of view, there is no way to prepare initial densities which are singular, so we shall focus on measures which are limits of transformations experienced by an initial smooth distribution $\rho(x)$ under the action of f ,

$$\rho_0(x) = \lim_{t \rightarrow \infty} \int_{\mathcal{M}} dy \delta(x - f^t(y)) \rho(y, 0), \quad \int_{\mathcal{M}} dy \rho(y, 0) = 1. \quad (19.15)$$

Intuitively, the “natural” measure should be the measure that is the least sensitive to the (in practice unavoidable) external noise, no matter how weak, or round-off errors in a numerical computation.

19.4.1 Natural measure

Huang: Chen-Ning, do you think ergodic theory gives us useful insight into the foundation of statistical mechanics?

Yang: I don’t think so.

—Kerson Huang, *C.N. Yang interview*



In computer experiments, as the Hénon example of figure 19.4, the long time evolution of many “typical” initial conditions leads to the same asymptotic distribution. Hence the *natural* measure (also called equilibrium measure, SRB measure, Sinai-Bowen-Ruelle measure, physical measure, invariant density, natural density, or even “natural invariant”) is defined as the limit

exercise 19.8
exercise 19.9

$$\bar{\rho}_{x_0}(y) = \begin{cases} \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t d\tau \delta(y - f^\tau(x_0)) & \text{flows} \\ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \delta(y - f^k(x_0)) & \text{maps,} \end{cases} \quad (19.16)$$

where x_0 is a generic initial point. Generated by the action of f , the natural measure satisfies the stationarity condition (19.14) and is thus invariant by construction.

Staring at an average over infinitely many Dirac deltas is not a prospect we cherish. From a computational point of view, the natural measure is the visitation frequency defined by coarse-graining, integrating (19.16) over the $\mathcal{M}_i$ region

$$\Delta \bar{\mu}_i = \lim_{t \rightarrow \infty} \frac{t_i}{t}, \quad (19.17)$$

where t_i is the accumulated time that a trajectory of total duration t spends in the $\mathcal{M}_i$ region, with the initial point x_0 picked from some smooth density $\rho(x)$.



Let $a = a(x)$ be any *observable*. In the mathematical literature $a(x)$ is a function belonging to some function space, for instance the space of integrable functions L^1 , that associates to each point in state space a number or a set of numbers. In physical applications the observable $a(x)$ is necessarily a smooth function. The observable reports on some property of the dynamical system. Several examples will be given in sect. 20.1.



The *space average* of the observable a with respect to a measure ρ is given by the d -dimensional integral over the state space $\mathcal{M}$:

$$\begin{aligned} \langle a \rangle_\rho &= \frac{1}{|\rho \mathcal{M}|} \int_{\mathcal{M}} dx \rho(x) a(x) \\ |\rho \mathcal{M}| &= \int_{\mathcal{M}} dx \rho(x) = \text{mass in } \mathcal{M}. \end{aligned} \quad (19.18)$$

For now we assume that the state space $\mathcal{M}$ has a finite dimension and a finite volume. By its construction, $\langle a \rangle_\rho$ is a function(al) of ρ . For $\rho = \rho_0$ natural measure we shall drop the subscript in the definition of the space average; $\langle a \rangle_\rho = \langle a \rangle$.

Inserting the right-hand-side of (19.16) into (19.18), we see that the natural measure corresponds to a *Birkhoff average* of the observable a along a trajectory of the initial point x_0 ,

$$\overline{a_{x_0}} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t d\tau a(f^\tau(x_0)). \quad (19.19)$$



Analysis of the above asymptotic time limit is the central problem of ergodic theory. The *Birkhoff ergodic theorem* asserts that if an invariant measure ρ exists, the limit $\overline{a(x_0)}$ for the time average (19.19) exists for (almost) all initial x_0 . Still, Birkhoff theorem says nothing about the dependence on x_0 of Birkhoff averages $\overline{a_{x_0}}$ (or, equivalently, that the construction of natural measures (19.16) leads to a “single” density, independent of x_0). This leads to one of the possible definitions of *ergodic* evolution: f is ergodic if for any integrable observable a in (19.19) the limit function is constant. If a flow enjoys such a property, the time averages coincide (apart from a set of ρ measure 0) with space averages

remark 19.1
appendix A1

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t d\tau a(f^\tau(x_0)) = \langle a \rangle. \quad (19.20)$$

For future reference, we note a further property that is stronger than ergodicity: if the space average of a product of any two variables decorrelates with time,

section 27.3

$$\lim_{t \rightarrow \infty} \langle a(x)b(f^t(x)) \rangle = \langle a \rangle \langle b \rangle, \quad (19.21)$$

the dynamical system is said to be *mixing*. The terminology may be understood better once we consider as the pair of observables in (19.21) the indicators of two sets $\mathcal{A}$ and $\mathcal{B}$: then (19.21) may be written as

$$\lim_{t \rightarrow \infty} \frac{\mu(\mathcal{A} \cap f^t(\mathcal{B}))}{\mu(\mathcal{A})} = \mu(\mathcal{B})$$

so that the set $\mathcal{B}$ spreads “uniformly” over the whole state space as t increases. Mixing is a fundamental notion in characterizing statistical behavior for dynamical systems: suppose we start with an arbitrary smooth nonequilibrium distribution $\rho(x)\nu(x)$: the after time t the average of an observable a is given by

$$\int_{\mathcal{M}} dx \rho(x) \nu(f^t(x)) a(x)$$

and this tends to the equilibrium average $\langle a \rangle_\rho$ if f is mixing.

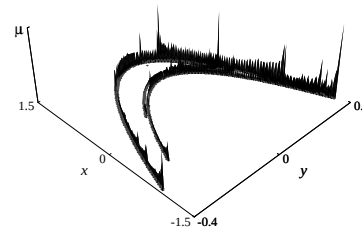


example 19.2
p. 398

remark A1.4

If an invariant measure is quite singular—for instance a Dirac δ concentrated on a fixed point or a cycle—it is most likely of no physical import. No smooth

Figure 19.4: Natural measure (19.17) for the Hénon map (3.18) strange attractor at parameter values $(a, b) = (1.4, 0.3)$. See figure 3.7 for a sketch of the attractor without the natural measure binning. See example 19.2. (Courtesy of J.-P. Eckmann)



initial density will converge to this measure if its neighborhood is repelling. In practice the average (19.16) is problematic and often hard to control, as generic dynamical systems are neither uniformly hyperbolic nor structurally stable: it is not known whether even the simplest model of a strange attractor, the Hénon attractor of figure 19.4, is “strange,” or merely a transient to a very long stable cycle.

exercise 6.3

19.4.2 Determinism vs. stochasticity



While dynamics can lead to very singular ρ 's, in any physical setting we cannot do better than to measure ρ averaged over some region M_i ; the coarse-graining is not an approximation but a physical necessity. One is free to think of a measure as a probability density, as long as one keeps in mind the distinction between deterministic and stochastic flows. In deterministic evolution the evolution kernels are not probabilistic; the density of trajectories is transported *deterministically*. What this distinction means will become apparent later: for deterministic flows our trace and determinant formulas will be *exact*, while for quantum and stochastic flows they will only be the leading saddle point (stationary phase, steepest descent) approximations.

chapter 22

Clearly, while deceptively easy to define, measures spell trouble. The good news is that if you hang on, you will *never need to compute them*, at least not in ChaosBook. How so? The evolution operators to which we next turn, and the trace and determinant formulas to which they will lead us, will assign the correct weights to desired averages without recourse to any explicit computation of the coarse-grained measure $\Delta\rho_i$.

19.5 Density evolution for infinitesimal times



Consider the evolution of a smooth density $\rho(x) = \rho(x, 0)$ under an infinitesimal step $\delta\tau$, by expanding the action of $\mathcal{L}^{\delta\tau}$ to linear order in $\delta\tau$:

$$\begin{aligned}\mathcal{L}^{\delta\tau}\rho(y) &= \int_{\mathcal{M}} dx \delta(y - f^{\delta\tau}(x)) \rho(x) \\ &= \int_{\mathcal{M}} dx \delta(y - x - \delta\tau v(x)) \rho(x) \\ &= \frac{\rho(y - \delta\tau v(y))}{\left| \det \left(1 + \delta\tau \frac{\partial v(y)}{\partial x} \right) \right|} = \frac{\rho(y) - \delta\tau v_i(y) \partial_i \rho(y)}{1 + \delta\tau \sum_{i=1}^d \partial_i v_i(y)} \\ \rho(x, \delta\tau) &= \rho(x, 0) - \delta\tau \frac{\partial}{\partial x} (v(x) \rho(x, 0)).\end{aligned}\tag{19.22}$$

Here we have used the infinitesimal form of the flow (2.9), the Dirac delta Jacobian (19.9), and the $\ln \det = \text{tr} \ln$ relation. By the Einstein summation convention, repeated indices imply summation, $v_i(y) \partial_i = \sum_{i=1}^d v_i(y) \partial_i$. Moving $\rho(y, 0)$ to the left hand side and dividing by $\delta\tau$, we discover that the rate of the deformation of ρ under the infinitesimal action of the Perron-Frobenius operator is nothing but the *continuity equation* for the density:

exercise 4.1

$$\partial_i \rho + \partial \cdot (\rho v) = 0.\tag{19.23}$$

From (19.22), time evolution by an infinitesimal step $\delta\tau$ forward in time is generated by

$$\mathcal{A}\rho(x) = + \lim_{\delta\tau \rightarrow 0^+} \frac{1}{\delta\tau} (\mathcal{L}^{\delta\tau} - I) \rho(x) = -\partial_i (v_i(x) \rho(x)).\tag{19.24}$$

We shall refer to

$$\mathcal{A} = -\partial \cdot v - \sum_i^d v_i(x) \partial_i\tag{19.25}$$

as the time-evolution *generator*. If the flow is finite-dimensional and invertible, $\mathcal{A}$ is a generator of a full-fledged group. The left hand side of (19.24) is the definition of time derivative, so the evolution equation for $\rho(x)$ is

$$\left(\frac{\partial}{\partial t} - \mathcal{A} \right) \rho(x) = 0.\tag{19.26}$$

appendix A31.2

The finite time Perron-Frobenius operator (19.10) can be formally expressed by exponentiating the time evolution generator $\mathcal{A}$ as

$$\mathcal{L}^t = e^{t\mathcal{A}}.\tag{19.27}$$

The generator $\mathcal{A}$ is reminiscent of the generator of translations. Indeed, for a constant velocity field dynamical evolution is nothing but a translation by (time $\times$ velocity):

$$e^{-tv \frac{\partial}{\partial x}} a(x) = a(x - tv).\tag{19.28}$$

exercise 19.10

19.6 Liouville operator



A case of special interest is the Hamiltonian or symplectic flow defined by Hamilton's equations of motion (8.1). A reader versed in quantum mechanics will have observed by now that with replacement $\mathcal{A} \rightarrow -\frac{i}{\hbar}\hat{H}$, where $\hat{H}$ is the quantum Hamiltonian operator, (19.26) looks rather like the time dependent Schrödinger equation, so this is the right moment to figure out what all this means for Hamiltonian flows.

The Hamilton's evolution equations (8.1) for any time-independent quantity $Q = Q(q, p)$ are given by

$$\frac{dQ}{dt} = \frac{\partial Q}{\partial q_i} \frac{dq_i}{dt} + \frac{\partial Q}{\partial p_i} \frac{dp_i}{dt} = \frac{\partial H}{\partial p_i} \frac{\partial Q}{\partial q_i} - \frac{\partial Q}{\partial p_i} \frac{\partial H}{\partial q_i}, \quad (19.29)$$

where (p_i, q_i) span the full state space, which for Hamiltonian flows we shall refer to as the *phase space*. As equations with this structure arise frequently for symplectic flows, it is convenient to introduce a notation for them, the *Poisson bracket*

remark 19.4

$$\{A, B\} = \frac{\partial A}{\partial p_i} \frac{\partial B}{\partial q_i} - \frac{\partial A}{\partial q_i} \frac{\partial B}{\partial p_i}. \quad (19.30)$$

In terms of Poisson brackets the time-evolution equation (19.29) takes the compact form

$$\frac{dQ}{dt} = \{H, Q\}. \quad (19.31)$$

The discussion of sect. 19.5 applies to any deterministic flow. The full phase space flow velocity is $\dot{x} = v = (\dot{q}, \dot{p})$, where the dot signifies time derivative.

section 33.1

If the density itself is a material invariant, combining

$$\partial_t I + v \cdot \partial I = 0.$$

and (19.23) we conclude that $\partial_i v_i = 0$ and $\det J^t(x_0) = 1$. An example of such incompressible flow is the Hamiltonian flow. For incompressible flows the continuity equation (19.23) becomes a statement of conservation of the phase space volume (see sect. 8.4), or the *Liouville theorem*

$$\partial_t \rho + v_i \partial_i \rho = 0. \quad (19.32)$$

The symplectic structure of Hamilton's equations (8.1) implies that the flow is incompressible, $\partial_i v_i = 0$, so for Hamiltonian flows the equation for ρ reduces to the *continuity equation* for the phase-space density:

$$\partial_t \rho + \partial_i(\rho v_i) = 0, \quad i = 1, 2, \dots, D. \quad (19.33)$$

Consider the evolution of the phase-space density ρ of an ensemble of noninteracting particles; the particles are conserved, so

$$\frac{d}{dt}\rho(q, p, t) = \left(\frac{\partial}{\partial t} + \dot{q}_i \frac{\partial}{\partial q_i} + \dot{p}_i \frac{\partial}{\partial p_i} \right) \rho(q, p, t) = 0.$$

Inserting Hamilton's equations (8.1) we obtain the *Liouville equation*, a special case of (19.26):

remark 19.4

$$\frac{\partial}{\partial t}\rho(q, p, t) = -\mathcal{A}\rho(q, p, t) = \{H, \rho(q, p, t)\}, \quad (19.34)$$

where $\{ , \}$ is the Poisson bracket (19.30). The generator of the flow (19.25) is in this case a generator of infinitesimal symplectic transformations,

$$\mathcal{A} = \dot{q}_i \frac{\partial}{\partial q_i} + \dot{p}_i \frac{\partial}{\partial p_i} = \frac{\partial H}{\partial p_i} \frac{\partial}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial}{\partial p_i}. \quad (19.35)$$

For example, for separable Hamiltonians of form $H = p^2/2m + V(q)$, the equations of motion are

$$\dot{q}_i = \frac{p_i}{m}, \quad \dot{p}_i = -\frac{\partial V(q)}{\partial q_i}. \quad (19.36)$$

and the action of the generator

exercise 19.11

$$\mathcal{A} = -\frac{p_i}{m} \frac{\partial}{\partial q_i} + \partial_i V(q) \frac{\partial}{\partial p_i}. \quad (19.37)$$

Looking back at (19.28) we see that the first term generates a translation in the configuration space, $f(q, p) \rightarrow f(q - dt \dot{q}, p)$, and the second generates acceleration by force $\partial V(q)$ in the momentum space. They do not commute, hence the time integration is not trivial.

The time-evolution generator (19.25) for the case of symplectic flows is called the *Liouville operator*. You might have encountered it in statistical mechanics, while discussing what ergodicity means for $6.02214129 \times 10^{23}$ hard balls. Here its action will be very tangible; we shall apply the Liouville operator to systems as small as 1 or 2 hard balls and to our surprise learn that this suffices to already get a bit of a grip on foundations of the nonequilibrium statistical mechanics.



in depth:

sect. A31.2, p. 1059

Résumé

In physically realistic settings the initial state of a system can be specified only to a finite precision. If the dynamics is chaotic, it is not possible to calculate the long time trajectory of a given initial point. Depending on the desired precision, and



given a deterministic law of evolution, the state of the system can then be tracked for a finite time only.

The study of long-time dynamics thus requires trading in the evolution of a single state space point for the evolution of a *measure*, or the *density* of representative points in state space, acted upon by an *evolution operator*. Essentially this means trading in *nonlinear* dynamical equations on a finite dimensional space $x = (x_1, x_2 \cdots x_d)$ for a *linear* equation on an infinite dimensional vector space of density functions $\rho(x)$. For finite times and for maps such densities are evolved by the *Perron-Frobenius operator*,

$$\rho(x, t) = (\mathcal{L}^t \circ \rho)(x),$$

and in a differential formulation they satisfy the *continuity equation*:

$$\partial_t \rho + \partial \cdot (\rho v) = 0.$$

The most physical of stationary measures is the natural measure, a measure robust under perturbations by weak noise.

Reformulated this way, deterministic dynamics takes on a distinctly quantum-mechanical flavor. If the Lyapunov time (1.1), the time after which the notion of an individual deterministic trajectory loses meaning, is much shorter than the observation time, the “sharp” observables are those dual to time, the eigenvalues of evolution operators. This is very much the same situation as in quantum mechanics; as atomic time scales are so short, what is measured is the energy, the quantum-mechanical observable dual to the time. Both in deterministic and quantum mechanics one has a choice of implementing dynamical evolution on densities (“Schrödinger picture,” sect. 19.5) or on observables (“Heisenberg picture,” sect. 20.2 and chapter 21).

In what follows we shall find the second formulation more convenient, but the alternative is worth keeping in mind when posing and solving invariant density problems. However, as deterministic evolution operators are not unitary, their eigenfunctions can be quite singular and difficult to work with. In what follows we shall learn how to avoid dealing with these eigenstates altogether. As a matter of fact, what follows will be a labor of radical deconstruction; after having argued so strenuously here that only smooth measures are “natural,” we shall merrily proceed to erect the whole edifice of our theory on periodic orbits, i.e., objects that are δ -functions in state space. The trick is that each comes with an interval, its neighborhood – periodic points only serve to pin these intervals, just as millimeter markings on a measuring rod are used to partition a continuum into intervals.

Commentary

Remark 19.1. Ergodic theory: An overview of ergodic theory is outside the scope of this book: the interested reader may find it useful to consult refs. [2, 17, 22, 26]. The existence of time average (19.19) is the basic result of ergodic theory, known as the

Birkhoff theorem, see for example refs. [16, 26], or the statement of theorem 7.3.1 in ref. [19]. The natural measure (19.17) of sect. 19.4.1 is often referred to as the SRB or Sinai-Ruelle-Bowen measure [5, 24, 25]. If you experience discomfort whenever a Dirac function is trotted out, *Ten Lessons*, Gian-Carlo Rota [23] sensible discussion of ‘density functions’ should bring you peace (“From this definition, all properties of the Dirac delta function are easily derived without any hysterical appeals to functions taking infinite values ...”).

There is much literature on explicit form of natural measure for special classes of 1-dimensional maps [3, 8, 20] - J. M. Aguirregabiria [1], for example, discusses several families of maps with known smooth measure, and behavior of measure under smooth conjugacies. As no such explicit formulas exist for higher dimensions and general dynamical systems, we do not discuss such measures here.

Remark 19.2. Time evolution as a Lie group: Time evolution of sect. 19.5 is an example of a 1-parameter Lie group. Consult, for example, Bluman and Kumei [4] Chapter 2 for a clear and pedagogical introduction to Lie groups of transformations. For a discussion of the bounded semigroups of page 410 see, for example, Marsden and Hughes [21].

Remark 19.3. Discretization of the Perron-Frobenius operator operator It is an old idea of Ulam [27] that such an approximation for the Perron-Frobenius operator is a meaningful one. The piecewise-linear approximation of the Perron-Frobenius operator (19.12) has been shown to reproduce the spectrum for expanding maps, once finer and finer Markov partitions are used [7, 9, 12]. The subtle point of choosing a state space partitioning for a “generic case” is discussed in ref. [10, 11].

Remark 19.4. The sign convention of the Poisson bracket: The Poisson bracket is antisymmetric in its arguments and there is a freedom to define it with either sign convention. When such freedom exists, it is certain that both conventions are in use and this is no exception. In some texts [13, 14] you will see the right hand side of (19.30) defined as $\{B, A\}$ so that (19.31) is $\frac{dQ}{dt} = \{Q, H\}$. Other equally reputable texts [15] employ the convention used here. Landau and Lifshitz [18] denote a Poisson bracket by $[A, B]$, notation that we reserve here for the quantum-mechanical commutator. As long as one is consistent, there should be no problem.

Remark 19.5. “Anon it lives”? “Anon it lives” refers to a statue of King Leontes’s wife, Hermione, who died in a fit of grief after he unjustly accused her of infidelity. Twenty years later, the servant Paulina shows Leontes this statue of Hermione. When he repents, the statue comes to life. Or perhaps Hermione actually lived and Paulina has kept her hidden all these years. The text of the play seems deliberately ambiguous. It is probably a parable for the resurrection of Christ.

(John F. Gibson)

References

- [1] J. M. Aguirregabiria, “Robust chaos with variable Lyapunov exponent in smooth one-dimensional maps”, *Chaos Solit. Fract.* **42**, 2531–2539 (2009).
- [2] V. I. Arnol’d and A. Avez, *Ergodic Problems of Classical Mechanics* (Addison-Wesley, Redwood City, 1989).

- [3] L. Billings and E. M. Bollt, “Invariant densities for skew tent maps”, *Chaos Solit. Fract.* **12**, 365 (2001).
- [4] G. W. Bluman and S. Kumei, *Symmetries and Differential Equations* (Springer, New York, 1989).
- [5] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms* (Springer, New York, 1975).
- [6] P. Cvitanović, C. P. Dettmann, R. Mainieri, and G. Vattay, “Trace formulas for stochastic evolution operators: Weak noise perturbation theory”, *J. Stat. Phys.* **93**, 981–999 (1998).
- [7] M. Dellnitz, G. Froyland, and S. Sertl, “On the isolated spectrum of the Perron-Frobenius operator”, *Nonlinearity* **13**, 1171 (2000).
- [8] D. J. Driebe, *Fully Chaotic Maps and Broken Time Symmetry* (Springer, New York, 1999).
- [9] G. Froyland, “Computer-assisted bounds for the rate of decay of correlations”, *Commun. Math. Phys.* **189**, 237–257 (1997).
- [10] G. Froyland, “Using Ulam’s method to calculate entropy and other dynamical invariants”, *Nonlinearity* **12**, 79–101 (1999).
- [11] G. Froyland, Extracting dynamical behaviour via Markov models, in *Nonlinear Dynamics and Statistics: Proc. Newton Institute, Cambridge 1998*, edited by A. Mees (2001), pp. 281–321.
- [12] G. Froyland, “On Ulam approximation of the isolated spectrum and eigenfunctions of hyperbolic maps”, *Discrete Contin. Dyn. Syst.* **17**, 671–689 (2007).
- [13] P. Gaspard, *Chaos, Scattering and Statistical Mechanics* (Cambridge Univ. Press, Cambridge UK, 1998).
- [14] H. Goldstein, *Classical Mechanics*, 2nd ed. (Wesley, Reading, MA, 1980).
- [15] M. C. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer, New York, 1990).
- [16] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems* (Cambridge Univ. Press, Cambridge UK, 1995).
- [17] I. Kornfeld, S. Fomin, and Y. G. Sinai, *Ergodic Theory* (Springer, New York, 1982).
- [18] L. D. Landau and E. M. Lifshitz, *Mechanics*, 3rd ed. (Pergamon Press, Oxford, 1959).
- [19] A. Lasota and M. MacKey, *Chaos, Fractals, and Noise; Stochastic Aspects of Dynamics* (Springer, New York, 1994).
- [20] J.-R. Luévano and E. Pina, “The Schröder functional equation and its relation to the invariant measures of chaotic maps”, *J. Phys. A.* **41**, 265101 (2008).
- [21] J. E. Marsden and T. J. R. Hughes, *Mathematical Foundations of Elasticity* (Dover, New York, 1994).
- [22] M. Pollicott and M. Yuri, *Dynamical systems and ergodic theory* (Cambridge Univ. Press, Cambridge UK, 1998).

- [23] G.-C. Rota, [Ten lessons I wish I had learned before I started teaching differential equations](#), tech. rep., Simmons College Meeting of Math. Assoc. of America (Massachusetts Inst. of Technology, 1997).
- [24] D. Ruelle, [Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics](#), 2nd ed. (Cambridge Univ. Press, Cambridge UK, 2004).
- [25] Y. G. Sinai, “Gibbs measures in ergodic theory”, [Russ. Math. Surv.](#) **27**, 21–69 (1972).
- [26] Y. G. Sinai, [Introduction to Ergodic Theory](#) (Princeton Univ. Press, Princeton, 1976).
- [27] S. M. Ulam, *A Collection of Mathematical Problems* (Interscience, New York, 1960).

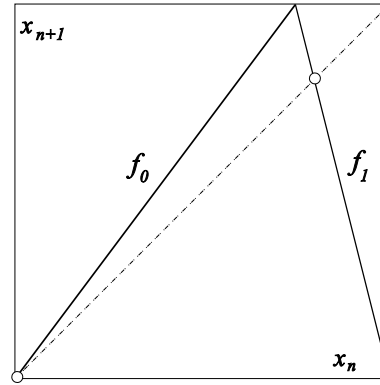


Figure 19.5: The piecewise-linear skew full tent map (19.38), with $\Lambda_0 = 4/3$, $\Lambda_1 = -4$. See example 19.1.

19.7 Examples

Example 19.1. Perron-Frobenius operator for a piecewise-linear map. Consider the expanding 1-dimensional map $f(x)$ of figure 19.5, a piecewise-linear 2-branch map with slopes $\Lambda_0 > 1$ and $\Lambda_1 = -\Lambda_0/(\Lambda_0 - 1) < -1$:

$$f(x) = \begin{cases} f_0(x) = \Lambda_0 x, & x \in \mathcal{M}_0 = [0, 1/\Lambda_0) \\ f_1(x) = \Lambda_1(1 - x), & x \in \mathcal{M}_1 = (1/\Lambda_0, 1] \end{cases} \quad (19.38)$$

exercise 19.7

Both $f(\mathcal{M}_0)$ and $f(\mathcal{M}_1)$ map onto the entire unit interval $\mathcal{M} = [0, 1]$. We shall refer to any unimodal map whose critical point maps onto the “left” unstable fixed point x_0 as the “Ulam” map. Assume a piecewise constant density

$$\rho(x) = \begin{cases} \rho_0 & \text{if } x \in \mathcal{M}_0 \\ \rho_1 & \text{if } x \in \mathcal{M}_1 \end{cases} \quad (19.39)$$

As can be easily checked using (19.9), the Perron-Frobenius operator acts on this piecewise constant function as a $[2 \times 2]$ Markov matrix (transfer matrix) $\mathbf{L}$ with matrix elements

$$\begin{bmatrix} \rho_0 \\ \rho_1 \end{bmatrix} \rightarrow \mathbf{L}\rho = \begin{bmatrix} \frac{1}{|\Lambda_0|} & \frac{1}{|\Lambda_1|} \\ \frac{1}{|\Lambda_0|} & \frac{1}{|\Lambda_1|} \end{bmatrix} \begin{bmatrix} \rho_0 \\ \rho_1 \end{bmatrix}, \quad (19.40)$$

exercise 19.1
exercise 19.5

stretching both ρ_0 and ρ_1 over the whole unit interval Λ . In this example the density is constant after one iteration, so $\mathbf{L}$ has only a unit eigenvalue $e^{s_0} = 1/|\Lambda_0| + 1/|\Lambda_1| = 1$, with constant density eigenvector $\rho_0 = \rho_1$. The quantities $1/|\Lambda_0|, 1/|\Lambda_1|$ are, respectively, the fractions of state space taken up by the $|\mathcal{M}_0|, |\mathcal{M}_1|$ intervals. This simple explicit matrix representation of the Perron-Frobenius operator is a consequence of the piecewise linearity of f , and the restriction of the densities ρ to the space of piecewise constant functions. The example gives a flavor of the enterprise upon which we are about to embark in ChaosBook, but the full story is much subtler: in general, there will exist no such finite-dimensional representation for the Perron-Frobenius operator. (continued in example 20.4)

return to p. 384

Example 19.2. The Hénon attractor natural measure. A numerical calculation of the natural measure (19.17) for the Hénon attractor (3.18) is given by the histogram in figure 19.4. The state space is partitioned into many equal-size areas $\mathcal{M}_i$, and the coarse grained measure (19.17) is computed by a long-time iteration of the Hénon map, and represented by the height of the column over area $\mathcal{M}_i$. What we see is a typical invariant measure - a complicated, singular function concentrated on a fractal set.

return to p. 389

19.8 Exercises

19.1. Integrating over Dirac delta functions.

Check the delta function integrals in

- (a) 1 dimension (19.7),

$$\int dx \delta(h(x)) = \sum_{\{x:h(x)=0\}} \frac{1}{|h'(x)|}, \quad (19.41)$$

- (b) and in d dimensions (19.8), $h : \mathbb{R}^d \rightarrow \mathbb{R}^d$,

$$\begin{aligned} \int_{\mathbb{R}^d} dx \delta(h(x)) &= \sum_j \int_{\mathcal{M}_j} dx \delta(h(x)) \\ &= \sum_{\{x:h(x)=0\}} \frac{1}{|\det \frac{\partial h(x)}{\partial x}|}. \end{aligned}$$

where $\mathcal{M}_j$ are arbitrarily small regions enclosing the zeros x_j (with x_j not on the boundary $\partial\mathcal{M}_j$). For a refresher on Jacobian determinants, read, for example, Stone and Goldbart Sect. 12.2.2.

- (c) The delta function can be approximated by a sequence of Gaussians

$$\int dx \delta(x) f(x) = \lim_{\sigma \rightarrow 0} \int dx \frac{e^{-\frac{x^2}{2\sigma}}}{\sqrt{2\pi\sigma}} f(x).$$

Use this approximation to see whether the formal expression

$$\int_{\mathbb{R}} dx \delta(x^2)$$

makes sense.

19.2. Derivatives of Dirac delta functions.

Consider $\delta^{(k)}(x) = \frac{\partial^k}{\partial x^k} \delta(x)$.

Using integration by parts, determine the value of

$$\begin{aligned} \int_{\mathbb{R}} dx \delta'(y) \quad , \quad \text{where } y = f(x) - x & \quad (19.42) \\ \int dx \delta^{(2)}(y) &= \sum_{\{x:y(x)=0\}} \frac{1}{|y'|} \left\{ 3 \frac{(y'')^2}{(y')^4} - \frac{y'''}{(y')^3} \right\} \\ \int dx b(x) \delta^{(2)}(y) &= \sum_{\{x:y(x)=0\}} \frac{1}{|y'|} \left\{ \frac{b''}{(y')^2} - \frac{b'y''}{(y')^3} \right. \\ &\quad \left. + b \left(3 \frac{(y'')^2}{(y')^4} - \frac{y'''}{(y')^3} \right) \right\}. \end{aligned}$$

These formulas are useful for computing effects of weak noise on deterministic dynamics [6].

- 19.3. **$\mathcal{L}^t$ generates a semigroup.** Check that the Perron-Frobenius operator has the semigroup property,

$$\int_M dz \mathcal{L}^{t_2}(y, z) \mathcal{L}^{t_1}(z, x) = \mathcal{L}^{t_2+t_1}(y, x), \quad t_1, t_2 \geq 0. \quad (19.43)$$

As the flows in which we tend to be interested are invertible, the $\mathcal{L}$'s that we will use often do form a group, with $t_1, t_2 \in \mathbb{R}$.

- 19.4. **Escape rate of the tent map.**

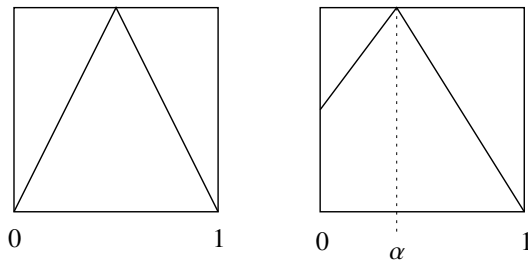
- (a) Calculate by numerical experimentation the log of the fraction of trajectories remaining trapped in the interval $[0, 1]$ for the tent map

$$f(x) = a(1 - 2|x - 0.5|)$$

for several values of a .

- (b) Determine analytically the a dependence of the escape rate $\gamma(a)$.
 (c) Compare your results for (a) and (b).

- 19.5. **Invariant measure.** We will compute the invariant measure for two different piecewise linear maps.



- (a) Verify the matrix $\mathcal{L}$ representation (19.40).
 (b) The maximum value of the first map is 1. Compute an invariant measure for this map.
 (c) Compute the leading eigenvalue of $\mathcal{L}$ for this map.
 (d) For this map there is an infinite number of invariant measures, but only one of them will be found when one carries out a numerical simulation. Determine that measure, and explain why your choice is the natural measure for this map.
 (e) In the second map the maximum occurs at $\alpha = (3 - \sqrt{5})/2$ and the slopes are $\pm(\sqrt{5} + 1)/2$. Find the natural measure for this map. Show that it is piecewise linear and that the ratio of its two values is $(\sqrt{5} + 1)/2$.

(medium difficulty)

- 19.6. **Escape rate for a flow conserving map.** Adjust Λ_0, Λ_1 in (19.38) so that the gap between the intervals $\mathcal{M}_0, \mathcal{M}_1$ vanishes. Show that the escape rate equals zero in this situation.

- 19.7. **Eigenvalues of the Perron-Frobenius operator for the skew full tent map.** Show that for the skew full tent map, see figure 19.5

$$f(x) = \begin{cases} f_0(x) = \Lambda_0 x, & x \in \mathcal{M}_0 = [0, 1/\Lambda_0) \\ f_1(x) = \frac{\Lambda_0}{\Lambda_0 - 1}(1 - x), & x \in \mathcal{M}_1 = (1/\Lambda_0, 1]. \end{cases} \quad (19.44)$$

the eigenvalues are available analytically, compute the first few.

19.8. **“Kissing disks”*** (continuation of exercises 9.1 and 9.2) Close off the escape by setting $R = 2$, and look in real time at the density of the Poincaré section iterates for a trajectory with a randomly chosen initial condition. Does it look uniform? Should it be uniform? (Hint - phase-space volumes are preserved for Hamiltonian flows by the Liouville theorem). Do you notice the trajectories that loiter near special regions of phase space for long times? These exemplify “intermittency,” a bit of unpleasantness to which we shall return in chapter 29.

19.9. **Invariant measure for the Gauss map.** Consider the Gauss map:

$$f(x) = \begin{cases} \frac{1}{x} - \left[\frac{1}{x} \right] & x \neq 0 \\ 0 & x = 0 \end{cases} \quad (19.45)$$

where $[\]$ denotes the integer part.

(a) Verify that the density

$$\rho(x) = \frac{1}{\log 2} \frac{1}{1+x}$$

is an invariant measure for the map.

(b) Is it the natural measure?

19.10. **$\mathcal{A}$ as a generator of translations.** Verify that for a constant velocity field the evolution generator $\mathcal{A}$ in (19.28) is the generator of translations,

$$e^{tv \frac{\partial}{\partial x}} a(x) = a(x + tv).$$

A hint: you have heard of a thing called ‘Taylor expansion’, no?

19.11. **Incompressible flows.** Show that (19.9) implies that $\rho_0(x) = 1$ is an eigenfunction of a volume-preserving flow with eigenvalue $s_0 = 0$. In particular, this implies that the natural measure of hyperbolic and mixing Hamiltonian flows is uniform. Compare this results with the numerical experiment of exercise 19.8.

Chapter 20

Averaging

Why think when you can compute?

—Maciej Zworski

WE DISCUSS FIRST the necessity of studying the averages of observables in chaotic dynamics. A time average of an observable is computed by integrating its value along a trajectory. The integral along trajectory can be split into a sum of over integrals evaluated on trajectory segments; if the observable is exponentiated, this yields a *multiplicative* weight for successive trajectory segments. This elementary observation will enable us to recast the formulas for averages in a multiplicative form that motivates the introduction of evolution operators and further formal developments to come. The main result is that any *dynamical* average measurable in a chaotic system can be extracted from the spectrum of an appropriately constructed evolution operator. In order to keep our toes closer to the ground, in sect. 20.4 we try out the formalism on the first quantitative diagnosis whether a system is chaotic, the Lyapunov exponent.



20.1 Dynamical averaging



In chaotic dynamics detailed prediction is impossible, as any finitely specified initial condition, no matter how precise, will fill out the entire accessible state space after a finite Lyapunov time (1.1). Hence for chaotic dynamics one cannot follow individual trajectories for a long time; what is attainable, however, is a description of the geometry of the set of possible outcomes, and the evaluation of long-time averages. Examples of such averages are transport coefficients for chaotic dynamical flows, such as escape rates, mean drifts and diffusion rates; power spectra; and a host of mathematical constructs such as generalized dimensions, entropies, and Lyapunov exponents. Here we outline how such averages are evaluated within the evolution operator framework. The key idea is to replace the expectation values of observables by the expectation values of exponential generating functionals. This

associates an evolution operator with a given observable, and relates the expectation value of the observable to the leading eigenvalue of the evolution operator.



20.1.1 Time averages



Let $a = a(x)$ be any *observable*, a function that associates to each point in state space a number, a vector, or a tensor. The observable reports on a property of the dynamical system. The observable is a device, such as a thermometer or laser Doppler velocitometer. The device itself does not change during the measurement. The velocity field $a_i(x) = v_i(x)$ is an example of a vector observable; the speed $|v(x)|$ (the length of this vector), or perhaps a temperature measured in an experiment at instant τ are examples of scalar observable. We define the *Birkhoff sum* A as the time integral of the observable a evaluated along the trajectory of the initial point x_0 ,

$$A(x_0, t) = \int_0^t d\tau a(x(\tau)), \quad x(t) = f^t(x_0). \quad (20.1)$$

If the dynamics are given by an iterated mapping and the time is discrete, the Birkhoff sum after n iterations is given by

$$A(x_0, n) = \sum_{k=0}^{n-1} a(x_k), \quad x_k = f^k(x_0) \quad (20.2)$$

(we suppress vectorial indices for the time being).

remark 20.1



example 20.1
p. 415

The *Birkhoff average* of the observable along an orbit is defined by

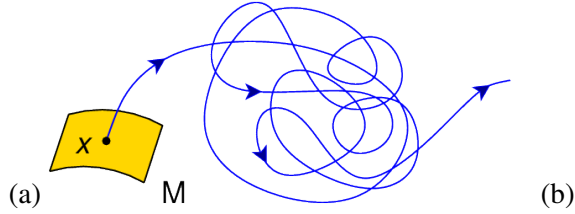
$$\overline{a(x_0)} = \lim_{t \rightarrow \infty} \frac{1}{t} A(x_0, t). \quad (20.3)$$

If a does not behave too wildly as a function of time—for example, if $a(x)$ is the Chicago temperature, bounded between $-80^\circ F$ and $+130^\circ F$ for all times— $A(x_0, t)$ is expected to grow no faster than t , and the limit (20.3) exists. For an example of a Birkhoff average—the Lyapunov exponent—see sect. 20.4.

The Birkhoff average is a property of the orbit, independent of the initial point on that orbit: if we start at a later state space point $f^T(x_0)$ we get a couple of extra finite contributions that vanish in the $t \rightarrow \infty$ limit:

$$\begin{aligned} \overline{a(f^T(x_0))} &= \lim_{t \rightarrow \infty} \frac{1}{t} \int_T^{t+T} d\tau a(f^\tau(x_0)) \\ &= \overline{a(x_0)} - \lim_{t \rightarrow \infty} \frac{1}{t} \left(\int_0^T d\tau a(f^\tau(x_0)) - \int_t^{t+T} d\tau a(f^\tau(x_0)) \right) \\ &= \overline{a(x_0)}. \end{aligned}$$

Figure 20.1: (a) A typical chaotic trajectory explores the state space with the long time visitation frequency building up the natural measure $\rho_0(x)$. (b) Time average, evaluated along an atypical trajectory such as a periodic orbit, fails to explore the entire accessible state space. (A. Johansen)



The Birkhoff sum $A(x_0, t)$ and the Birkhoff average $\overline{a(x_0)}$ take a particularly simple form when evaluated on a periodic orbit. Define

$$A_p = \begin{cases} a_p T_p & = \int_0^{T_p} d\tau a(x(\tau)) & \text{for a flow} \\ a_p n_p & = \sum_{i=1}^{n_p} a(x_i) & \text{for a map} \end{cases}, \quad x \in \mathcal{M}_p, \quad (20.4)$$

exercise 4.6

where p is a prime cycle, T_p is its period, and n_p is its discrete time period in the case of iterated map dynamics. The quantity A_p is a loop integral of the observable along a single traversal of a prime cycle p , so it is an intrinsic property of the cycle, independent of the starting point $x_0 \in \mathcal{M}_p$. If the trajectory retraces itself r times, we just obtain A_p repeated r times. Evaluation of the asymptotic Birkhoff average (20.3) therefore requires only a single traversal of the cycle:

$$a_p = A_p / T_p. \quad (20.5)$$

Innocent as this seems, it implies that $\overline{a(x_0)}$ is in general a wild function of x_0 ; for a hyperbolic system it takes the same value $\langle a \rangle$ for almost all initial x_0 , but a different value (20.5) on (almost) every periodic orbit (figure 20.1 (b)).



example 20.2
p. 415

section 24.1

20.1.2 Spatial averages



The *space average* of a quantity a evaluated over all state space trajectories $x(t)$ at time t is given by the d -dimensional integral over all initial points x_0 at time $t = 0$:

$$\begin{aligned} \langle a \rangle(t) &= \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} dx_0 a(x(t)), & x(t) &= f^t(x_0) \\ |\mathcal{M}| &= \int_{\mathcal{M}} dx = \text{volume of } \mathcal{M}. \end{aligned} \quad (20.6)$$

The space $\mathcal{M}$ is assumed to have finite volume - open systems like the 3-disk game of pinball are discussed in sect. 20.3.



example 20.6
p. 417

What is it we *really* do in experiments? We cannot measure the Birkhoff average (20.3), as there is no way to prepare a single initial condition with infinite precision. The best we can do is prepare an initial density $\rho(x)$, perhaps concentrated on some small (but always finite) neighborhood. Then we can abandon the uniform space average (20.6) and consider instead the weighted spatial average

$$\langle a \rangle_\rho(t) = \frac{1}{|\mathcal{M}_\rho|} \int_{\mathcal{M}} dx_0 \rho(x_0) a(x(t)), \quad |\mathcal{M}_\rho| = \int_{\mathcal{M}} dx \rho(x). \quad (20.7)$$

For ergodic mixing systems, *any* smooth initial density will tend to the asymptotic natural measure in the $t \rightarrow \infty$ limit $\rho(x, t) \rightarrow \rho_0(x)$. This allows us to take any smooth initial $\rho(x)$ and define the *expectation value* $\langle a \rangle$ of an observable a as the asymptotic time and space average over the state space $\mathcal{M}$

$$\langle a \rangle = \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} dx \overline{a(x)} = \lim_{t \rightarrow \infty} \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} dx_0 \frac{1}{t} \int_0^t d\tau a(x(t)). \quad (20.8)$$

We use the same $\langle \cdot \cdot \cdot \rangle$ notation as for the space average (20.6) and distinguish the two by the presence of the time variable in the argument: if the quantity $\langle a \rangle(t)$ being averaged depends on time, then it is a space average; if it is the infinite time limit, it is the expectation value $\langle a \rangle$.

The expectation value is a space average of Birkhoff averages, with every $x \in \mathcal{M}$ used as a starting point of a Birkhoff average. The advantage of averaging over space is that it smears the starting points which were problematic for the Birkhoff average (such as periodic points). While easy to define, the expectation value $\langle a \rangle$ turns out not to be particularly tractable in practice. 

Here comes a simple idea that is the basis of all that follows: Such averages are more conveniently studied by investigating instead of $\langle a \rangle$ the space averages of form 

$$\langle e^{\beta \cdot A} \rangle = \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} dx e^{\beta \cdot A(x, t)}. \quad (20.9)$$

In the present context β is an auxiliary variable of no physical significance whose role is to enable us to recover the desired space average by differentiation,

$$\langle A_i \rangle = \left. \frac{\partial}{\partial \beta_i} \langle e^{\beta \cdot A} \rangle \right|_{\beta=0}.$$

We write ' $\beta \cdot A$ ' to indicate that if the observable is a d -dimensional vector $a(x) \in \mathbb{R}^d$, then $\beta \in \mathbb{R}^d$; if the observable is a $[d \times d]$ tensor, β is also a rank-2 tensor, and so on. Here we will mostly limit the considerations to scalar β and drop the dot in ' $\beta \cdot A$ '.

If the Birkhoff average limit $a(x_0)$ (20.3) exists for 'almost all' initial x_0 's and the system is ergodic and mixing (in the sense of sect. 1.3.1), we expect the Birkhoff average along almost all trajectories to tend to the same value $\bar{a}$, and the Birkhoff sum A to tend to $t\bar{a}$. The space average (20.9) is an integral over exponentials and hence also grows (or shrinks) exponentially with time. So as

$t \rightarrow \infty$ we would expect the space average of $\exp(\beta A(x, t))$ to grow exponentially with time

$$\langle e^{\beta A} \rangle \rightarrow (\text{const}) e^{ts(\beta)},$$

and its rate of growth (or contraction) characteristic state function to be given by the limit

$$s(\beta) = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \langle e^{\beta A} \rangle. \quad (20.10)$$

Now we understand one reason for why it is smarter to compute $\langle \exp(\beta A) \rangle$ rather than $\langle a \rangle$: the expectation value of the observable (20.8), the (generalized) diffusion tensor, and higher moments of the Birkhoff sum (20.1) can be computed by evaluating the derivatives of $s(\beta)$

$$\begin{aligned} \left. \frac{\partial s}{\partial \beta_j} \right|_{\beta=0} &= \lim_{t \rightarrow \infty} \frac{1}{t} \langle A_j \rangle &= \langle a_j \rangle, \\ \left. \frac{\partial^2 s}{\partial \beta_i \partial \beta_j} \right|_{\beta=0} &= \lim_{t \rightarrow \infty} \frac{1}{t} \left(\langle A_i A_j \rangle - \langle A_i \rangle \langle A_j \rangle \right) \\ &= \lim_{t \rightarrow \infty} \frac{1}{t} \langle (A_i - t \langle a_i \rangle)(A_j - t \langle a_j \rangle) \rangle = \Delta_{ij}, \end{aligned} \quad (20.11)$$

and so forth. We have explicitly written out the formulas for a scalar observable; the vector case is worked out in exercise 20.1 (we could have used full derivative notation $ds/d\beta$ in (20.11), but for vector observable we do need partial derivatives $\partial s/\partial \beta_i$). If we can compute the function $s(\beta)$, we have the desired expectation value without having to estimate any infinite time limits from finite time data.

exercise 20.1

Suppose we could evaluate $s(\beta)$ and its derivatives. What are such formulas good for? A typical application arises in the problem of determining transport coefficients from underlying deterministic dynamics.



example 20.3
p. 415

We now turn to the problem of evaluating $\langle e^{\beta A} \rangle$, but first you might want to review some elementary notions of probability theory that will be useful later on.



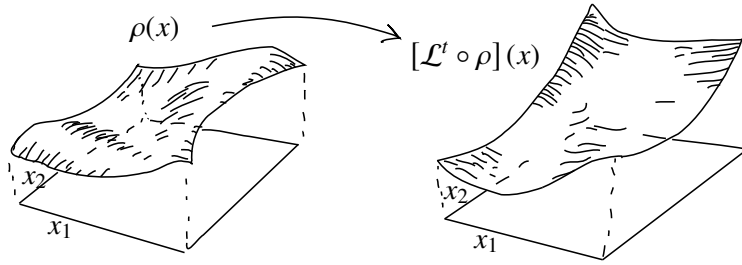
in depth:
sect. A20.1, p. 994

20.2 Evolution operators

For it, the mystic evolution;
Not the right only justified
– what we call evil also justified.

—Walt Whitman,
Leaves of Grass: Song of the Universal

Figure 20.2: Space averaging pieces together the Birkhoff average computed along the $t \rightarrow \infty$ orbit of figure 20.1 by a space average over infinitely many short t trajectory segments starting at all initial points at once.



The above simple shift of focus, from studying $\langle a \rangle$ to studying $\langle \exp(\beta A) \rangle$ is the key to everything that follows. Make the dependence on the flow explicit by rewriting this quantity as

$$\langle e^{\beta A} \rangle = \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} dx \int_{\mathcal{M}} dy \delta(y - f^t(x)) e^{\beta A(x,t)}. \quad (20.12)$$

Here $\delta(y - f^t(x))$ is the Dirac delta function: for a deterministic flow an initial point x maps into a unique point y at time t . Formally, all we have done above is to insert the identity

$$1 = \int_{\mathcal{M}} dy \delta(y - f^t(x)), \quad (20.13)$$

into (20.9) to make explicit the fact that we are averaging only over the trajectories that remain in $\mathcal{M}$ for all times. However, having made this substitution we have replaced the study of individual trajectories $f^t(x)$ by studying the evolution of the density of *the totality* of initial conditions. Instead of trying to extract a temporal average from an arbitrarily long trajectory which explores the state space ergodically, we can now probe the entire state space with short (and controllable) finite time pieces of trajectories originating from every point in $\mathcal{M}$.

As a matter of fact (and that is why we went to the trouble of defining the generator (19.25) of infinitesimal transformations of densities) *infinitesimally* short time evolution induced by the generator $\mathcal{A}$ of (19.25) suffices to determine the spectrum and eigenvalues of $\mathcal{L}^t$.



We shall refer to the kernel of the operation (20.12) as the *evolution operator*

$$\mathcal{L}^t(y, x) = \delta(y - f^t(x)) e^{\beta A(x,t)}. \quad (20.14)$$

The simplest example is the $\beta = 0$ case, i.e., the Perron-Frobenius operator introduced in sect. 19.2. Another example - designed to deliver the Lyapunov exponent - will be the evolution operator (20.33) discussed below. The action of the evolution operator on a function ϕ is given by

$$[\mathcal{L}^t \phi](y) = \int_{\mathcal{M}} dx \delta(y - f^t(x)) e^{\beta A(x,t)} \phi(x). \quad (20.15)$$

The evolution operator is different for different observables, as its definition depends on the choice of the Birkhoff sum A in the exponential. Its job is to deliver

the expectation value of a , but before showing that it accomplishes that, we need to verify the semigroup property of evolution operators. 

By its definition, the integral over the observable a is additive along the trajectory

$$\begin{aligned}
 & \begin{array}{c} \text{Diagram: A trajectory from } x(0) \text{ to } x(t_1+t_2) \text{ via } x(t_1). \end{array} \\
 A(x_0, t_1 + t_2) &= \int_0^{t_1+t_2} d\tau a(f^\tau(x)) \\
 &= \int_0^{t_1} d\tau a(f^\tau(x)) + \int_{t_1}^{t_1+t_2} d\tau a(f^\tau(x)) \\
 &= A(x_0, t_1) + A(f^{t_1}(x_0), t_2).
 \end{aligned}$$

As $A(x, t)$ is additive along the trajectory, the evolution operator generates a semigroup

exercise 19.3
section 19.5

$$\mathcal{L}^{t_1+t_2}(y, x) = \int_{\mathcal{M}} dz \mathcal{L}^{t_2}(y, z) \mathcal{L}^{t_1}(z, x), \quad (20.16)$$

as is easily checked by substitution

$$[\mathcal{L}^{t_2} \mathcal{L}^{t_1} a](y) = \int_{\mathcal{M}} dx \delta(y - f^{t_2}(x)) e^{\beta A(x, t_2)} [\mathcal{L}^{t_1} a](x) = [\mathcal{L}^{t_1+t_2} a](y).$$

This semigroup property is the main reason why (20.12) is preferable to (20.8) as a starting point for evaluation of dynamical averages: it recasts averaging in form of operators multiplicative along the flow.

In terms of the evolution operator, the space average of the moment-generating function (20.12) is given by

$$\langle e^{\beta A} \rangle = \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} dx \int_{\mathcal{M}} dy \phi(y) \mathcal{L}^t(y, x) \phi(x).$$

where $\phi(x)$ is the constant function $\phi(x) = 1$. If the linear operator $\mathcal{L}^t$ can be thought of as a matrix, high powers of a matrix are dominated by its fastest growing matrix elements, and the limit (20.10)

$$s(\beta) = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \langle \mathcal{L}^t \rangle. \quad (20.17)$$

yields the leading eigenvalue $s_0(\beta)$, and, through it, all desired expectation values (20.11).

In what follows we shall learn how to extract not only the leading eigenvalue of $\mathcal{L}^t$, but much of the dominant part of its spectrum. Clearly, we are not interested into the eigenvalues of $\mathcal{L}^t$ for any particular finite time t , but their behavior as $t \rightarrow \infty$. That is achieved *via* a Laplace transform, see sect. 20.2.3.

20.2.1 Spectrum of an evolution operator

This operator is strange:
it is not self-adjoint, so it is nothing good
—Jean Bellissard



An exposition of a subject is of necessity sequential and one cannot explain everything at once. As we shall actually never use eigenfunctions of evolution operators, we postpone their discussion to sect. 28.6. For the time being we ask the reader to accept uncritically the following sketch:



Schematically, a linear operator has a spectrum of eigenvalues s_α and eigenfunctions $\varphi_\alpha(x)$

$$[\mathcal{L}^t \varphi_\alpha](x) = e^{s_\alpha t} \varphi_\alpha(x), \quad \alpha = 0, 1, 2, \dots \quad (20.18)$$

ordered so that $\text{Re } s_\alpha \geq \text{Re } s_{\alpha+1}$. For continuous time flow eigenvalues cannot depend on time, they are eigenvalues of the time-evolution generator (19.24) we always write the eigenvalues of an evolution operator in exponentiated form e^{s_α} rather than as multipliers λ_α . We find it convenient to write them this way both for the continuous time $\mathcal{L}^t$ and the discrete time $\mathcal{L} = \mathcal{L}^1$ cases, and we shall assume that spectrum of $\mathcal{L}$ is discrete.



$\mathcal{L}^t$ is a linear operator acting on a density of initial conditions $\rho(x)$, $x \in \mathcal{M}$, so the $t \rightarrow \infty$ limit will be dominated by $s_0 = s(\beta)$, the leading eigenvalue of $\mathcal{L}^t$,

$$[\mathcal{L}^t \rho_\beta](y) := \int_{\mathcal{M}} dx \delta(y - f^t(x)) e^{\beta A(x,t)} \rho_\beta(x) = e^{ts(\beta)} \rho_\beta(y), \quad (20.19)$$

where $\rho_\beta(x)$ is the corresponding eigenfunction. For $\beta = 0$ the evolution operator (20.14) is the Perron-Frobenius operator (19.10), with $\rho_0(x)$ the natural measure.



From now on we have to be careful to distinguish the two kinds of linear operators. In chapter 5 we have characterized the evolution of the *local* linear neighborhood of a state space trajectory by eigenvalues and eigenvalues of the linearized flow Jacobian matrices. Evolution operators described in this chapter are *global*, and they act on densities of orbits, not on individual trajectories. As we shall see, one of the wonders of chaotic dynamics is that the more unstable individual trajectories, the nicer are the corresponding global density functions.

20.2.2 Evolution for infinitesimal times

For infinitesimal time δt , the evolution operator (20.6) acts as

$$\begin{aligned} \rho(y, \delta t) &= \int dx e^{\beta A(x, \delta t)} \delta(y - f^{\delta t}(x)) \rho(x, 0) \\ &= \int dx e^{\beta a(x) \delta t} \delta(y - x - \delta t v(x)) \rho(x, 0) \\ &= (1 + \delta t \beta a(y)) \frac{\rho(y, 0) - \delta t v \cdot \frac{\partial}{\partial x} \rho(y, 0)}{1 + \delta t \frac{\partial v}{\partial x}}, \end{aligned}$$

(the denominator arises from the δt linearization of the jacobian) giving the continuity equation (19.23) a source term

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_i} (v_i \rho) = \beta a \rho. \quad (20.20)$$

The evolution generator (19.25) eigenfunctions now satisfy

$$(s(\beta) - \mathcal{A})\rho(x, \beta) = \beta a(x)\rho(x, \beta). \quad (20.21)$$

Differentiating with respect to β

$$\begin{aligned} s'(\beta)\rho(x, \beta) + s(\beta)\frac{\partial}{\partial\beta}\rho(x, \beta) + \frac{\partial}{\partial x}\left(v(x)\frac{\partial}{\partial\beta}\rho(x, \beta)\right) \\ = a(x)\rho(x, \beta) + \beta a(x)\frac{\partial}{\partial\beta}\rho(x, \beta) \end{aligned}$$

In the vanishing auxiliary parameter limit $\beta \rightarrow 0$, we have $s(0) = 0$, $\rho(x, 0) = \rho_0(x)$

$$s'(0)\rho_0(x) + \frac{\partial}{\partial x_i}\left(v_i(x)\frac{\partial}{\partial\beta}\rho(x, 0)\right) = a(x)\rho_0(x).$$

By integrating, the second term vanishes by Gauss' theorem

$$s'(0) = \int dx a(x)\rho_0(x) = \langle a \rangle,$$

verifying equation (20.7): spatial average of the observable a is given by the derivative of the leading eigenvalue $s'(0)$.



fast track:
sect. 21, p. 419

20.2.3 Resolvent of $\mathcal{L}$

Here we limit ourselves to a brief remark about the notion of the ‘spectrum’ of a linear operator.

The Perron-Frobenius operator $\mathcal{L}$ acts multiplicatively in time, so it is reasonable to suppose that there exist constants $M > 0$, $s_0 \geq 0$ such that $\|\mathcal{L}^t\| \leq Me^{ts_0}$ for all $t \geq 0$. What does that mean? The operator norm is defined in the same spirit in which one defines matrix norms: We are assuming that no value of $\mathcal{L}^t\rho(x)$ grows faster than exponentially for any choice of function $\rho(x)$, so that the fastest possible growth can be bounded by e^{ts_0} , a reasonable expectation in the light of the simplest example studied so far, the escape rate (1.3). If that is so, multiplying $\mathcal{L}^t$ by e^{-ts_0} we construct a new operator $e^{-ts_0}\mathcal{L}^t = e^{t(\mathcal{A}-s_0I)}$ which decays exponentially for large t , $\|e^{t(\mathcal{A}-s_0I)}\| \leq M$. We say that $e^{-ts_0}\mathcal{L}^t$ is an element of a *bounded* semigroup with generator $\mathcal{A} - s_0I$. Given this bound, it follows by the Laplace transform

$$\int_0^\infty dt e^{-st}\mathcal{L}^t = \frac{1}{s - \mathcal{A}}, \quad \text{Re } s > s_0, \quad (20.22)$$

remark A22.1

that the *resolvent* operator $(s - \mathcal{A})^{-1}$ is bounded

$$\left\| \frac{1}{s - \mathcal{A}} \right\| \leq \int_0^\infty dt e^{-st} M e^{ts_0} = \frac{M}{s - s_0}. \quad (20.23)$$

If one is interested in the spectrum of $\mathcal{L}$, as we will be, the resolvent operator is a natural object to study; it has no time dependence, and it is bounded. It is called ‘resolvent’ because it separates the spectrum of $\mathcal{L}$ into individual constituents, one for each spectral ‘line’. From (20.17), it is clear that the leading eigenvalue $s_0(\beta)$ corresponds to the pole in (20.23); as we shall see in chapter 21, the rest of the spectrum is similarly resolved into further poles of the Laplace transform.



The main lesson of this brief aside is that for continuous time flows, the Laplace transform is the tool that brings down the generator in (19.27) into the resolvent form (20.22) and enables us to study its spectrum.



in depth:
appendix A31.2, p. 1059

20.3 Averaging in open systems



If $\mathcal{M}$ is a compact region or set of regions to which the dynamics is confined for all times, (20.8) is a sensible definition of the expectation value. However, if the trajectories can exit $\mathcal{M}$ without ever returning,

$$\int_{\mathcal{M}} dy \delta(y - f^t(x_0)) = 0 \quad \text{for } t > t_{\text{exit}}, \quad x_0 \in \mathcal{M},$$

we might be in trouble. In particular, a *repeller* is a dynamical system for which the trajectory $f^t(x_0)$ eventually leaves the region $\mathcal{M}$, unless the initial point x_0 is on the repeller, so the identity

$$\int_{\mathcal{M}} dy \delta(y - f^t(x_0)) = 1, \quad t > 0, \quad \text{iff } x_0 \in \text{non-wandering set} \quad (20.24)$$

might apply only to a fractal subset of initial points of zero Lebesgue measure (non-wandering set is defined in sect. 2.1.1). Clearly, for open systems we need to modify the definition of the expectation value to restrict it to the dynamics on the non-wandering set, the set of trajectories which are confined for all times.

Denote by $\mathcal{M}$ a state space region that encloses all interesting initial points, say the 3-disk Poincaré section constructed from the disk boundaries and all possible incidence angles, and denote by $|\mathcal{M}|$ the volume of $\mathcal{M}$. The volume of state space containing all trajectories, which start out within the state space region $\mathcal{M}$ and recur within that region at time t , is given by

$$|\mathcal{M}(t)| = \int_{\mathcal{M}} dx dy \delta(y - f^t(x)) \sim |\mathcal{M}| e^{-\gamma t}. \quad (20.25)$$

As we have already seen in sect. 1.4.3, this volume is expected to decrease exponentially, with the escape rate γ . The integral over x takes care of all possible initial points; the integral over y checks whether their trajectories are still within $\mathcal{M}$ by the time t . For example, any trajectory that falls off the pinball table in

section 27.1

figure 1.1 is gone for good.

If we expand an initial distribution $\rho(x)$ in (20.18), the eigenfunction basis $\rho(x) = \sum_{\alpha} a_{\alpha} \varphi_{\alpha}(x)$, we can also understand the rate of convergence of finite-time estimates to the asymptotic escape rate. For an open system the fraction of trapped trajectories decays as

section 20.3

$$\begin{aligned} \Gamma_{\mathcal{M}}(t) &= \frac{\int_{\mathcal{M}} dx [\mathcal{L}^t \rho](x)}{\int_{\mathcal{M}} dx \rho(x)} = \sum_{\alpha} e^{s_{\alpha} t} a_{\alpha} \frac{\int_{\mathcal{M}} dx \varphi_{\alpha}(x)}{\int_{\mathcal{M}} dx \rho(x)} \\ &= e^{s_0 t} \left((\text{const.}) + O(e^{(s_1 - s_0)t}) \right). \end{aligned} \quad (20.26)$$

The constant depends on the initial density $\rho(x)$ and the geometry of state space cutoff region $\mathcal{M}$, but the escape rate $\gamma = -s_0$ is an intrinsic property of the repelling set. We see, at least heuristically, that the leading eigenvalue of $\mathcal{L}^t$ dominates $\Gamma_{\mathcal{M}}(t)$ and yields the escape rate, a measurable property of a given repeller.

The non-wandering set can be very difficult to describe; but for any finite time we can construct a normalized measure from the finite-time covering volume (20.25), by redefining the space average (20.9) as

$$\langle e^{\beta A} \rangle = \int_{\mathcal{M}} dx \frac{1}{|\mathcal{M}(t)|} e^{\beta A(x,t)} \sim \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} dx e^{\beta A(x,t) + \gamma t}. \quad (20.27)$$

in order to compensate for the exponential decrease of the number of surviving trajectories in an open system with the exponentially growing factor $e^{\gamma t}$. What does this mean? Once we have computed γ we can replenish the density lost to escaping trajectories, by pumping in $e^{\gamma t}$ of new trajectories in such a way that the overall measure is correctly normalized at all times, $\langle 1 \rangle = 1$.

 example 20.4
p. 416

20.4 Evolution operator evaluation of Lyapunov exponents

A solution to these problems was offered in sect. 20.2: replace time averaging along a single orbit by action of a multiplicative evolution operator on the entire state space, and extract the state space average of the Lyapunov exponent from its leading eigenvalue, computed from finite period cycles. The main idea - what is the Lyapunov ‘observable’ - can be illustrated by the dynamics of a 1-dimensional map.

 example 20.5
p. 416

Here we have restricted our considerations to 1- d maps, as for higher-dimensional flows only the Jacobian matrices are multiplicative, not the individual eigenvalues. Construction of the evolution operator for evaluation of the Lyapunov spectra for

a d -dimensional flow requires more skill than warranted at this stage in the narrative: an extension of the evolution equations to a flow in the tangent space.

If the chaotic motion fills the whole state space, we are indeed computing the asymptotic Lyapunov exponent. If the chaotic motion is transient, leading eventually to some long attractive cycle, our Lyapunov exponent, computed on a non-wandering set, will characterize the chaotic transient; this is actually what any experiment would measure, as even a very small amount of external noise suffices to destabilize a long stable cycle with a minute immediate basin of attraction.

All that remains is to determine the value of the Lyapunov exponent

$$\lambda = \langle \ln |f'(x)| \rangle = \left. \frac{\partial s(\beta)}{\partial \beta} \right|_{\beta=0} = s'(0) \quad (20.28)$$

from (20.11), the derivative of the leading eigenvalue $s_0(\beta)$ of the evolution operator (20.33).

example 23.3

The only question is: How? (By chapter 23 you will know.)

Résumé

The expectation value $\langle a \rangle$ of an observable $a(x)$ integrated, $A^t(x) = \int_0^t d\tau a(x(\tau))$, and time averaged, A^t/t , over the trajectory $x \rightarrow x(t)$ is given by the derivative

$$\langle a \rangle = \left. \frac{\partial s}{\partial \beta} \right|_{\beta=0}$$

of the leading eigenvalue $e^{ts(\beta)}$ of the evolution operator $\mathcal{L}^t$.

By computing the leading eigenfunction of the Perron-Frobenius operator (19.10), one obtains the expectation value (19.18) of any observable $a(x)$. Thus we can construct a specific, hand-tailored evolution operator $\mathcal{L}$ for each and every observable. The good news is that, by the time we arrive at chapter 23, the scaffolding will be removed, both $\mathcal{L}$'s and their eigenfunctions will be gone, and only the explicit and exact periodic orbit formulas for expectation values of observables will remain.

chapter 23

The next question is: How do we evaluate the eigenvalues of $\mathcal{L}$? In example 20.4, we saw a piecewise-linear example where these operators reduce to finite matrices $\mathbf{L}$, but for generic smooth flows, they are infinite-dimensional linear operators, and finding smart ways of computing their eigenvalues requires some thought. In chapter 14 we undertook the first step, and replaced the *ad hoc* partitioning (19.12) by the intrinsic, topologically invariant partitioning. In chapter 18 we applied this information to our first application of the evolution operator formalism, evaluation of the topological entropy, and the growth rate of the number of topologically distinct orbits. In chapters 21 and 22, this small victory will be refashioned into a systematic method for computing eigenvalues of evolution operators in terms of periodic orbits.

Commentary

Remark 20.1. Birkhoff sum. In number theory, ergodic theory and dynamical systems literature the integrated observable (20.2) is called a ‘Birkhoff sum’, and the time average along an orbit (20.3) is called a ‘Birkhoff average’ [4].

Remark 20.2. ‘Pressure’. The quantity $\langle \exp(\beta A) \rangle$ in (20.9) is called a ‘partition function’ by Ruelle [9]. Some authors decorate it with considerably more Greek and Gothic letters than is done in this treatise. By a somewhat unfortunate analogy Ruelle [6] had given name ‘pressure’ or ‘topological pressure $P(a)$ ’ to $s(\beta)$ (where a is the observable introduced in sect. 20.1.1), defined by the ‘large system’ limit (20.10). In Ruelle’s words [10, 11] “I introduced the term ‘pressure’ for a function that appears in ergodic theory; frankly the proper term in statistical mechanics should be ‘free energy’, but the word pressure seemed more acceptable to mathematicians.” A more standard statistical mechanics name for $s(\beta)$ is ‘characteristic state function’. As the periodic orbit theory developed here can be applied to computing the physical gas pressure exerted on the walls of a container by a bouncing particle, we refer to $s(\beta)$ as simply the ‘leading eigenvalue’ of the evolution operator introduced in sect. 19.5. The ‘convexity’ properties such as $P(a) \leq P(|a|)$ will be pretty obvious consequences of the definition (20.10). In the case that $\mathcal{L}$ is the Perron-Frobenius operator (19.10), the eigenvalues $\{s_0(\beta), s_1(\beta), \dots\}$ are called the *Ruelle-Pollicott resonances* [5, 7, 8], with the leading one, $s(\beta) = s_0(\beta)$ being the one of main physical interest.

Remark 20.3. State space discretization. Ref. [1] discusses numerical discretizations of state space, and construction of Perron-Frobenius operators as stochastic matrices, or directed weighted graphs, as coarse-grained models of the global dynamics, with transport rates between state space partitions computed using this matrix of transition probabilities; a rigorous discussion of some of the former features is included in ref. [2].

References

- [1] M. Dellnitz, O. Junge, W. S. Koon, F. Lekien, M. W. Lo, J. E. Marsden, K. Padberg, R. Preis, S. D. Ross, and B. Thiere, “Transport in dynamical astronomy and multibody problems”, *Int. J. Bifur. Chaos* **15**, 699–727 (2005).
- [2] G. Froyland, “Computer-assisted bounds for the rate of decay of correlations”, *Commun. Math. Phys.* **189**, 237–257 (1997).
- [3] P. Gaspard and S. A. Rice, “Scattering from a classically chaotic repeller”, *J. Chem. Phys.* **90**, 2225–2241 (1989).
- [4] O. Knill and F. Tangerman, “Self-similarity and growth in Birkhoff sums for the golden rotation”, *Nonlinearity* **24**, 3115–3127 (2011).
- [5] M. Pollicott, “On the rate of mixing of Axiom A flows”, *Inv. Math.* **81**, 413–426 (1985).
- [6] D. Ruelle, “Statistical mechanics on a compact set with Z' action satisfying expansiveness and specification”, *Bull. Amer. Math. Soc.* **78**, 988–991 (1972).

- [7] D. Ruelle, “One-dimensional Gibbs states and Axiom A diffeomorphisms”, *J. Diff. Geom.* **25**, 117–137 (1987).
- [8] D. Ruelle, “Resonances for Axiom A flows”, *J. Diff. Geom.* **25**, 99–116 (1987).
- [9] D. Ruelle, *Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics*, 2nd ed. (Cambridge Univ. Press, Cambridge UK, 2004).
- [10] H. H. Rugh, “Un interview de David Ruelle”, *Gaz. Math.* **161**, 55–59 (2019).
- [11] H. H. Rugh, “An interview with David Ruelle”, *EMS Newsletter* **2020-3**, 21–24 (2020).

20.5 Examples

Example 20.1. Integrated observables.

(a) If the observable is the velocity, $a_i(x) = v_i(x)$, its time integral $A(x_0, t_i)$ is the trajectory $A(x_0, t_i) = x_i(t)$.

(b) For Hamiltonian flows the *action* associated with a trajectory $x(t) = [q(t), p(t)]$ passing through a phase-space point $x_0 = [q(0), p(0)]$ is the Birkhoff sum

$$A(x_0, t) = \int_0^t d\tau \dot{\mathbf{q}}(\tau) \cdot \mathbf{p}(\tau) \quad (20.29)$$

[return to p. 403](#)

Example 20.2. Deterministic diffusion. The phase space of an open system such as the Sinai gas (an infinite 2-dimensional periodic array of scattering disks, see sect. 24.1) is dense with initial points that correspond to periodic runaway trajectories. The mean distance squared traversed by any such trajectory grows as $x(t)^2 \sim t^2$, and its contribution to the diffusion rate $D \propto x(t)^2/t$, (20.3) evaluated with $a(x) = x(t)^2$, diverges. Seemingly there is a paradox; even though intuition says the typical motion should be diffusive, we have an infinity of ballistic trajectories.

For chaotic dynamical systems, this paradox is resolved by also averaging over the initial x and worrying about the measure of the ‘pathological’ trajectories. (continued in example 20.3)

[return to p. 404](#)

Example 20.3. Deterministic diffusion. (continued from example 20.2) Consider a point particle scattering elastically off a d -dimensional array of scatterers. If the scatterers are sufficiently large to block any infinite length free flights, the particle will diffuse chaotically, and the transport coefficient of interest is the diffusion constant $\langle x(t)^2 \rangle \approx 2dDt$. In contrast to D estimated numerically from trajectories $x(t)$ for finite but large t , the above formulas yield the asymptotic D without any extrapolations to the $t \rightarrow \infty$ limit. For example, for $a_i = v_i$ and zero mean drift $\langle v_i \rangle = 0$, in d dimensions the diffusion constant is given by the curvature of $s(\beta)$ at $\beta = 0$,

[section 24.1](#)

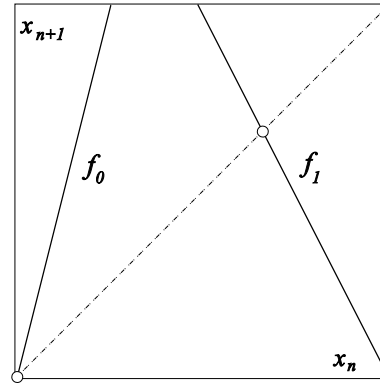


Figure 20.3: A piecewise-linear repeller (19.38): All trajectories that land in the gap between the f_0 and f_1 branches escape ($\Lambda_0 = 4$, $\Lambda_1 = -2$). See example 20.4.

$$D = \lim_{t \rightarrow \infty} \frac{1}{2dt} \langle x(t)^2 \rangle = \frac{1}{2d} \sum_{i=1}^d \left. \frac{\partial^2 s}{\partial \beta_i^2} \right|_{\beta=0}, \quad (20.30)$$

so if we can evaluate derivatives of $s(\beta)$, we can compute transport coefficients that characterize deterministic diffusion. As we shall see in chapter 24, periodic orbit theory yields an *exact* and explicit closed form expression for D .

[return to p. 406](#)

Example 20.4. Escape rate for a piecewise-linear repeller: (continuation of example 19.1) What is gained by reformulating the dynamics in terms of ‘operators’? We start by considering a simple example in which the operator is a $[2 \times 2]$ matrix. Assume the expanding 1-dimensional map $f(x)$ of figure 20.3, a piecewise-linear 2-branch repeller (19.38). Assume a piecewise constant density (19.39). There is no need to define $\rho(x)$ in the gap between M_0 and M_1 , as any point that lands in the gap escapes.

The physical motivation for studying this kind of mapping is the pinball game: f is the simplest model for the pinball escape, figure 1.8, with f_0 and f_1 modelling its two strips of survivors.

As can be easily checked using (19.9), the Perron-Frobenius operator acts on this piecewise constant function as a $[2 \times 2]$ ‘transfer’ matrix (Markov matrix) (19.40)

[exercise 19.1](#)
[exercise 19.5](#)

$$\begin{pmatrix} \rho_0 \\ \rho_1 \end{pmatrix} \rightarrow \mathcal{L}\rho = \begin{bmatrix} \frac{1}{|\Lambda_0|} & \frac{1}{|\Lambda_1|} \\ \frac{1}{|\Lambda_0|} & \frac{1}{|\Lambda_1|} \end{bmatrix} \begin{pmatrix} \rho_0 \\ \rho_1 \end{pmatrix},$$

stretching both ρ_0 and ρ_1 over the whole unit interval Λ , and decreasing the density at every iteration. In this example the density is constant after one iteration, so $\mathcal{L}$ has only one non-zero eigenvalue $e^{s_0} = 1/|\Lambda_0| + 1/|\Lambda_1| \leq 1$, with constant density eigenvector $\rho_0 = \rho_1$. The quantities $1/|\Lambda_0|$, $1/|\Lambda_1|$ are, respectively, the sizes of the $|M_0|$, $|M_1|$ intervals, so the *exact* escape rate (1.3) – the log of the fraction of survivors at each iteration for this linear repeller – is given by the sole eigenvalue of $\mathcal{L}$:

$$\gamma = -s_0 = -\ln(1/|\Lambda_0| + 1/|\Lambda_1|). \quad (20.31)$$

Voila! Here is the rationale for introducing operators – in one time step we have solved the problem of evaluating escape rates at infinite time. (continued in example 28.2)

[return to p. 412](#)

Example 20.5. Lyapunov exponent, discrete time 1-dimensional dynamics. Due to the chain rule (4.22) for the derivative of an iterated map, the stability of a 1-dimensional mapping is multiplicative along the flow, so the integral (20.1) of the observable $a(x) = \ln |f'(x)|$, the local trajectory divergence rate, evaluated along the trajectory of x_0 ,

is additive:

$$A(x_0, n) = \ln |f^{n'}(x_0)| = \sum_{k=0}^{n-1} \ln |f'(x_k)|. \quad (20.32)$$

The associated one time step evolution operator (20.14) is

$$\mathcal{L}(y, x) = \delta(y - f(x)) e^{\beta \ln |f'(x)|}. \quad (20.33)$$

For a 1-dimensional iterative mapping, the Lyapunov exponent is then the expectation value (20.8) given by a spatial integral (20.7) weighted by the natural measure

$$\lambda = \langle \ln |f'(x)| \rangle = \int_{\mathcal{M}} dx \rho_0(x) \ln |f'(x)|. \quad (20.34)$$

[return to p. 412](#)

Example 20.6. Microcanonical ensemble. In statistical mechanics the space average (20.6) performed over the Hamiltonian system constant energy surface invariant measure $\rho(x)dx = dqdp \delta(H(q, p) - E)$ of volume $\omega(E) = \int_{\mathcal{M}} dqdp \delta(H(q, p) - E)$

$$\langle a(t) \rangle = \frac{1}{\omega(E)} \int_{\mathcal{M}} dqdp \delta(H(q, p) - E) a(q, p; t) \quad (20.35)$$

is called the *microcanonical ensemble average*.

[return to p. 404](#)

Exercises

20.1. Expectation value of a vector observable.

Check and extend the expectation value formulas (20.11) by evaluating the derivatives of $s(\beta)$ up to 4-th order for the space average $\langle \exp(\beta \cdot A) \rangle$ with a_i a vector quantity:

(a)

$$\left. \frac{\partial s}{\partial \beta_i} \right|_{\beta=0} = \lim_{t \rightarrow \infty} \frac{1}{t} \langle A_i \rangle = \langle a_i \rangle, \quad (20.36)$$

(b)

$$\begin{aligned} \left. \frac{\partial^2 s}{\partial \beta_k \partial \beta_j} \right|_{\beta=0} &= \lim_{t \rightarrow \infty} \frac{1}{t} \left(\langle A_k A_j \rangle - \langle A_k \rangle \langle A_j \rangle \right) \\ &= \lim_{t \rightarrow \infty} \frac{1}{t} \langle (A_k - t \langle a_k \rangle)(A_j - t \langle a_j \rangle) \rangle. \end{aligned}$$

Note that the formalism is smart: it automatically yields the *variance* from the mean, rather than simply the 2nd moment $\langle a^2 \rangle$.

(c) compute the third derivative of $s(\beta)$.

(d) compute the fourth derivative assuming that the mean in (20.36) vanishes, $\langle a_i \rangle = 0$. The 4-th order moment formula

$$K(t) = \frac{\langle x^4(t) \rangle}{\langle x^2(t) \rangle^2} - 3 \quad (20.37)$$

that you have derived is known as *kurtosis* (A20.11): it measures a deviation from what the 4-th order moment would be were the distribution a pure Gaussian (see (24.14) for a concrete example). If the observable is a vector, the kurtosis $K(t)$ is given by

$$\frac{\sum_{kj} \left[\langle A_k A_k A_j A_j \rangle + 2 \left(\langle A_k A_j \rangle \langle A_j A_k \rangle - \langle A_k A_k \rangle \langle A_j A_j \rangle \right) \right]}{(\sum_k \langle A_k A_k \rangle)^2}$$

20.2. Escape rate for a 1-dimensional repeller, numerically.

Consider the quadratic map

$$f(x) = Ax(1 - x) \quad (20.38)$$

on the unit interval. The trajectory of a point starting in the unit interval either stays in the interval forever or after some iterate leaves the interval and diverges to minus infinity. Estimate numerically the escape rate (27.8), the rate of exponential decay of the measure of points remaining in the unit interval, for either $A = 9/2$ or $A = 6$. Remember to compare your numerical estimate with the solution of the continuation of this exercise, exercise 23.2.

20.3. Pinball escape rate from numerical simulation*.

Estimate the escape rate for $R : a = 6$ 3-disk pinball by shooting 100,000 randomly initiated pinballs into the 3-disk system and plotting the logarithm of the number of trapped orbits as function of time. For comparison, a numerical simulation of ref. [3] yields $\gamma = .410 \dots$

Chapter 21

Trace formulas

The trace formula is not a formula, it is an idea.

—Martin Gutzwiller

DYNAMICS IS POSED in terms of local equations, but the ergodic averages require global information. How can we use a local description of a flow to learn something about the global behavior? In chapter 20 we have related global averages to the eigenvalues of appropriate evolution operators. Here we show that the traces of evolution operators can be evaluated as integrals over Dirac delta functions, and in this way the spectra of evolution operators become related to periodic orbits. If there is one idea that one should learn about chaotic dynamics, it happens in this chapter, and it is this: there is a fundamental local $\leftrightarrow$ global duality which says that



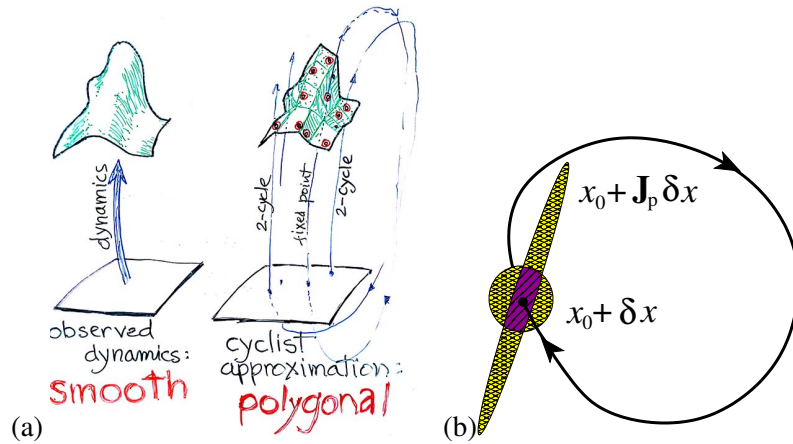
the spectrum of eigenvalues is dual to the spectrum of periodic orbits

For dynamics on the circle, this is called Fourier analysis; for dynamics on well-tiled manifolds, Selberg traces and zetas; and for generic nonlinear dynamical systems the duality is embodied in the trace formulas that we will now derive. These objects are to dynamics what partition functions are to statistical mechanics.

The above phrasing is a bit too highfalutin, so it perhaps pays to go again through the quick sketch of sects. 1.5 and 1.6. We have a state space that we would like to tessellate by periodic orbits, one short orbit per neighborhood, as in figure 21.1 (a). How big is the neighborhood of a given cycle?

Along stable directions neighbors of the periodic orbit get closer with time, so we only have to keep track of those who are moving away along the unstable directions. The fraction of those who remain in the neighborhood for one recurrence time T_p is given by the overlap ratio along the initial sphere and the returning ellipsoid, figure 21.1 (b), and along the expanding eigen-direction $\mathbf{e}^{(i)}$ of $J_p(x)$ this

Figure 21.1: (a) Smooth dynamics tessellated by the skeleton of periodic points, together with their linearized neighborhoods. (b) Jacobian matrix J_p maps spherical neighborhood of $x_0 \rightarrow$ ellipsoidal neighborhood time T_p later, with the overlap ratio along the expanding eigendirection $\mathbf{e}^{(i)}$ of $J_p(x)$ given by the expanding eigenvalue $1/|\Lambda_{p,i}|$.



is given by the expanding Floquet multiplier $1/|\Lambda_{p,i}|$. A bit more thinking leads to the conclusion that one also cares about how long it takes to return (the long returns contributing less to the time averages), so the weight t_p of the p -neighborhood $t_p = e^{-sT_p}/|\Lambda_p|$ decreases exponentially both with the shortest recurrence period and the product (5.7) of expanding Floquet multipliers $\Lambda_p = \prod_e \Lambda_{p,e}$. With emphasis on *expanding* - the flow could be a 60,000-dimensional dissipative flow, and still the neighborhood is defined by the handful of expanding eigen-directions. Now the long-time average of a physical observable -let us say power D dissipated by viscous friction of a fluid flowing through a pipe- can be estimated by its mean value (20.5) D_p/T_p computed on each neighborhood, and weighted by the above estimate

$$\langle D \rangle \approx \sum_p \frac{D_p}{T_p} \frac{e^{-sT_p}}{|\Lambda_p|}.$$

Wrong in detail, this estimate is the crux of many a *Phys. Rev. Letter*, and in its essence the key result of this chapter, the ‘trace formula.’ Here we redo the argument in a bit greater depth, and derive the correct formula (23.23) for a long time average $\langle D \rangle$ as a weighted sum over periodic orbits. It will take three chapters, but it is worth it - the reward is an *exact* (i.e., not heuristic) and highly convergent and controllable formula for computing averages over chaotic flows.

21.1 A trace formula for maps

Our extraction of the spectrum of $\mathcal{L}$ commences with the evaluation of the trace. As the case of discrete time mappings is somewhat simpler, we first derive the trace formula for maps, and then, in sect. 21.2, for flows. The final formula (21.19) covers both cases.

To compute an expectation value using (20.12) we have to integrate over all the values of the kernel $\mathcal{L}^n(x, y)$. Were $\mathcal{L}^n$ a matrix sum over its matrix elements would be dominated by the leading eigenvalue as $n \rightarrow \infty$ (we went through the argument in some detail in sect. 18.1). As the trace of $\mathcal{L}^n$ is also dominated by the

leading eigenvalue as $n \rightarrow \infty$, we might just as well look at the trace for which we have a very explicit formula

exercise 18.3

$$\mathrm{tr} \mathcal{L}^n = \int dx \mathcal{L}^n(x, x) = \int dx \delta(x - f^n(x)) e^{\beta A(x, n)}. \quad (21.1)$$

On the other hand, by its matrix motivated definition, the trace is the sum over eigenvalues (20.18),

$$\mathrm{tr} \mathcal{L}^n = \sum_{\alpha=0}^{\infty} e^{s_{\alpha} n}. \quad (21.2)$$

We find it convenient to write the eigenvalues as exponents $e^{s_{\alpha}}$ rather than as multipliers λ_{α} , and we assume that spectrum of $\mathcal{L}$ is discrete, $s_0, s_1, s_2, \dots$, ordered so that $\mathrm{Re} s_{\alpha} \geq \mathrm{Re} s_{\alpha+1}$.

For the time being we choose not to worry about convergence of such sums, ignore the question of what function space the eigenfunctions belong to, and compute the eigenvalue spectrum without constructing any explicit eigenfunctions. We shall revisit these issues in more depth in chapter 28, and discuss how lack of hyperbolicity leads to continuous spectra in chapter 29.

21.1.1 Hyperbolicity assumption



We have learned in sect. 19.2 how to evaluate the delta-function integral (21.1).

section 19.2

According to (19.8) the trace (21.1) picks up a contribution whenever $x - f^n(x) = 0$, i.e., whenever x belongs to a periodic orbit. For reasons which we will explain in sect. 21.2, it is wisest to start by focusing on discrete time systems. The contribution of an isolated prime cycle p of period n_p for a map f can be evaluated by restricting the integration to an infinitesimal open neighborhood $\mathcal{M}_p$ around the cycle,



$$\begin{aligned} \mathrm{tr}_p \mathcal{L}^{n_p} &= \int_{\mathcal{M}_p} dx \delta(x - f^{n_p}(x)) \\ &= \frac{n_p}{|\det(\mathbf{1} - M_p)|} = n_p \prod_{i=1}^d \frac{1}{|1 - \Lambda_{p,i}|}. \end{aligned} \quad (21.3)$$

For the time being we set here and in (19.9) the observable $e^{\beta A_p} = 1$. Periodic orbit Jacobian matrix M_p is also known as the *monodromy matrix*, and its eigenvalues $\Lambda_{p,1}, \Lambda_{p,2}, \dots, \Lambda_{p,d}$ as the Floquet multipliers.

section 5.2.1

We sort the eigenvalues $\Lambda_{p,1}, \Lambda_{p,2}, \dots, \Lambda_{p,d}$ of the p -cycle $[d \times d]$ monodromy matrix M_p into expanding, marginal and contracting sets $\{e, m, c\}$, as in (5.6). As the integral (21.3) can be evaluated only if M_p has no eigenvalue of unit magnitude, we assume that no eigenvalue is marginal (we shall show in sect. 21.2 that

the longitudinal $\Lambda_{p,d+1} = 1$ eigenvalue for flows can be eliminated by restricting the consideration to the transverse monodromy matrix M_p , and factorize the trace (21.3) into a product over the expanding and the contracting eigenvalues

$$\left| \det(\mathbf{1} - M_p) \right|^{-1} = \frac{1}{|\Lambda_p|} \prod_e \frac{1}{1 - 1/\Lambda_{p,e}} \prod_c \frac{1}{1 - \Lambda_{p,c}}, \quad (21.4)$$

where $\Lambda_p = \prod_e \Lambda_{p,e}$ is the product of expanding eigenvalues. Both $\Lambda_{p,c}$ and $1/\Lambda_{p,e}$ are smaller than 1 in absolute value, and as they are either real or come in complex conjugate pairs we are allowed to drop the absolute value brackets $|\cdots|$ in the above products.

The *hyperbolicity assumption* requires that the stabilities of all cycles included in the trace sums be exponentially bounded away from unity:

$$\begin{aligned} |\Lambda_{p,e}| &> e^{\lambda_e T_p} && \text{any } p, \text{ any expanding } |\Lambda_{p,e}| > 1 \\ |\Lambda_{p,c}| &< e^{-\lambda_c T_p} && \text{any } p, \text{ any contracting } |\Lambda_{p,c}| < 1, \end{aligned} \quad (21.5)$$

where $\lambda_e, \lambda_c > 0$ are strictly positive bounds on the expanding, contracting cycle Lyapunov exponents. If a dynamical system satisfies the hyperbolicity assumption (for example, the well separated 3-disk system clearly does), the $\mathcal{L}^t$ spectrum will be relatively easy to control. If the expansion/contraction is slower than exponential, let us say $|\Lambda_{p,i}| \sim T_p^2$, the system may exhibit “phase transitions,” and the analysis is much harder - we shall discuss this in chapter 29.



example 21.1
p. 430

It follows from (21.4) that for long times, $t = rT_p \rightarrow \infty$, only the product of expanding eigenvalues matters, $\left| \det(\mathbf{1} - M_p^r) \right| \rightarrow |\Lambda_p|^r$. We shall use this fact to motivate the construction of dynamical zeta functions in sect. 22.3. However, for evaluation of the full spectrum the exact cycle weight (21.3) has to be kept.

21.1.2 A deterministic trace formula for maps

If the evolution is given by a discrete time mapping, and all periodic points have Floquet multipliers $|\Lambda_{p,i}| \neq 1$ strictly bounded away from unity, the trace $\mathcal{L}^n$ is given by the sum over all *periodic points* i of period n :

$$\text{tr } \mathcal{L}^n = \int dx \mathcal{L}^n(x, x) = \sum_{x_i \in \text{Fix } f^n} \frac{e^{\beta A_i}}{|\det(\mathbf{1} - M^n(x_i))|}. \quad (21.6)$$

Here $\text{Fix } f^n = \{x : f^n(x) = x\}$ is the set of all periodic points of period n , and A_i is the observable (20.4) evaluated over n discrete time steps along the cycle to which the periodic point x_i belongs. The weight follows from the properties of the Dirac delta function (19.8) by taking the determinant of $\partial_i(x_j - f^n(x)_j)$. If a trajectory retraces itself r times, its monodromy matrix is M_p^r , where M_p is the

$[d \times d]$ monodromy matrix (4.5) evaluated along a single traversal of the prime cycle p . As we saw in (20.4), the Birkhoff sum A is additive along the cycle: If a prime cycle p trajectory retraces itself r times, $n = rn_p$, we obtain A_p repeated r times, $A_i = A(x_i, n) = rA_p$, $x_i \in \mathcal{M}_p$.

A prime cycle is a single traversal of the orbit, and its label is a non-repeating symbol string. There is only one prime cycle for each cyclic permutation class. For example, the four periodic points $\overline{0011} = \overline{1001} = \overline{1100} = \overline{0110}$ belong to the same prime cycle $p = 0011$ of period 4. As both the stability of a cycle and the weight A_p are the same everywhere along the orbit, each prime cycle of period n_p contributes n_p terms to the sum, one for each periodic point. Hence (21.6) can be rewritten as a sum over all prime cycles and their repeats

chapter 14

$$\text{tr } \mathcal{L}^n = \sum_p n_p \sum_{r=1}^{\infty} \frac{e^{r\beta A_p}}{|\det(\mathbf{1} - M_p^r)|} \delta_{n, n_p r}, \quad (21.7)$$

with the Kronecker delta $\delta_{n, n_p r}$ projecting out the periodic contributions of total period n . This constraint is awkward, and will be more awkward still for the continuous time flows, where it would yield a series of Dirac delta spikes. In both cases a Laplace transform rids us of the time periodicity constraint.

In the sum over all cycle periods,

$$\sum_{n=1}^{\infty} z^n \text{tr } \mathcal{L}^n = \text{tr } \frac{z\mathcal{L}}{1 - z\mathcal{L}} = \sum_p n_p \sum_{r=1}^{\infty} \frac{z^{n_p r} e^{r\beta A_p}}{|\det(\mathbf{1} - M_p^r)|}, \quad (21.8)$$

the constraint $\delta_{n, n_p r}$ is replaced by weight z^n . Such discrete time Laplace transform of $\text{tr } \mathcal{L}^n$ is usually referred to as a ‘generating function’. Why this transform? We are actually not interested in evaluating the sum (21.7) for any particular fixed period n ; what we are interested in is the long time $n \rightarrow \infty$ behavior. The transform trades in the large time n behavior for the small z behavior. Expressing the trace as in (21.2), in terms of the sum of the eigenvalues of $\mathcal{L}$, we obtain the *trace formula for maps*:



$$\sum_{\alpha=0}^{\infty} \frac{ze^{s_{\alpha}}}{1 - ze^{s_{\alpha}}} = \sum_p n_p \sum_{r=1}^{\infty} \frac{z^{n_p r} e^{r\beta A_p}}{|\det(\mathbf{1} - M_p^r)|}. \quad (21.9)$$

This is our second example of the duality between the spectrum of eigenvalues and the spectrum of periodic orbits, announced in the introduction to this chapter. (The first example was the topological trace formula (18.8).)



fast track:
sect. 21.2, p. 423



example 21.2
p. 430

21.2 A trace formula for flows

Amazing! I did not understand a single word.
—Fritz Haake

(R. Artuso and P. Cvitanović)

Our extraction of the spectrum of $\mathcal{L}^t$ commences with the evaluation of the trace

$$\mathrm{tr} \mathcal{L}^t = \mathrm{tr} e^{\mathcal{A}t} = \int dx \mathcal{L}^t(x, x) = \int dx \delta(x - f^t(x)) e^{\beta A(x, t)}. \quad (21.10)$$

We are not interested in any particular time t , but into the long-time behavior as $t \rightarrow \infty$, so we need to transform the trace from the “time domain” into the “frequency domain.” A generic flow is a semi-flow defined forward in time, so the appropriate transform is a Laplace rather than Fourier.

For a continuous time flow, the Laplace transform of an evolution operator yields the resolvent (20.22). This is a delicate step, since the evolution operator becomes the identity in the $t \rightarrow 0^+$ limit. In order to make sense of the trace we regularize the Laplace transform by a lower cutoff ϵ smaller than the period of any periodic orbit, and write

$$\int_{\epsilon}^{\infty} dt e^{-st} \mathrm{tr} \mathcal{L}^t = \mathrm{tr} \frac{e^{-(s-\mathcal{A})\epsilon}}{s - \mathcal{A}} = \sum_{\alpha=0}^{\infty} \frac{e^{-(s-s_{\alpha})\epsilon}}{s - s_{\alpha}}, \quad (21.11)$$

where $\mathcal{A}$ is the generator of the semigroup of dynamical evolution, see sect. 19.5. Our task is to evaluate $\mathrm{tr} \mathcal{L}^t$ from its explicit state space representation.

21.2.1 Integration along the flow

As any pair of nearby points on a cycle returns to itself exactly at each cycle period, the eigenvalue of the Jacobian matrix corresponding to the eigenvector along the flow necessarily equals unity for all periodic orbits. Hence for flows the trace integral $\mathrm{tr} \mathcal{L}^t$ requires a separate treatment for the longitudinal direction. To evaluate the contribution of an isolated prime cycle p of period T_p , restrict the integration to an infinitesimally thin tube $\mathcal{M}_p$ enveloping the cycle (see figure 1.12), and consider a local coordinate system with a longitudinal coordinate $dx_{\parallel}$ along the direction of the flow, and $d-1$ transverse coordinates $x_{\perp}$,

section 5.3.1

$$\mathrm{tr}_p \mathcal{L}^t = \int_{\mathcal{M}_p} dx_{\perp} dx_{\parallel} \delta(x_{\perp} - f_{\perp}^t(x)) \delta(x_{\parallel} - f_{\parallel}^t(x_{\parallel})). \quad (21.12)$$

(we set $\beta = 0$ in the $\exp(\beta \cdot A)$ weight for the time being). Pick a point on the prime cycle p , and let

$$v(x_{\parallel}) = \left(\sum_{i=1}^d v_i(x)^2 \right)^{1/2} \quad (21.13)$$

be the magnitude of the tangential velocity at any point $x = (x_{\parallel}, 0, \dots, 0)$ on the cycle p . The velocity $v(x)$ must be strictly positive, as otherwise the orbit would stagnate for infinite time at $v(x) = 0$ points, and that would get us nowhere.

As $0 \leq \tau < T_p$, the trajectory $x_{\parallel}(\tau) = f^{\tau}(x_p)$ sweeps out the entire cycle, and for larger times $x_{\parallel}$ is a cyclic variable of periodicity T_p ,

$$x_{\parallel}(\tau) = x_{\parallel}(\tau + rT_p) \quad r = 1, 2, \dots \quad (21.14)$$

We parametrize both the longitudinal coordinate $x_{||}(\tau)$ and the velocity $v(\tau) = v(x_{||}(\tau))$ by the flight time τ , and rewrite the integral along the periodic orbit as

$$\oint_p dx_{||} \delta(x_{||} - f^t(x_{||})) = \oint_p d\tau v(\tau) \delta(x_{||}(\tau) - x_{||}(\tau + t)). \quad (21.15)$$

By the periodicity condition (21.14) the Dirac δ function picks up contributions for $t = rT_p$, so the Laplace transform can be split as

$$\begin{aligned} \int_0^\infty dt e^{-st} \delta(x_{||}(\tau) - x_{||}(\tau + t)) &= \sum_{r=1}^\infty e^{-sT_p r} I_r \\ I_r &= \int_{-\epsilon}^\epsilon dt e^{-st} \delta(x_{||}(\tau) - x_{||}(\tau + rT_p + t)). \end{aligned}$$

Taylor expanding and applying the periodicity condition (21.14), we have $x_{||}(\tau + rT_p + t) = x_{||}(\tau) + v(\tau)t + \dots$,

$$I_r = \int_{-\epsilon}^\epsilon dt e^{-st} \delta(x_{||}(\tau) - x_{||}(\tau + rT_p + t)) = \frac{1}{v(\tau)},$$

so the remaining integral (21.15) over τ is simply the cycle period $\oint_p d\tau = T_p$. The contribution of the longitudinal integral to the Laplace transform is thus

$$\int_0^\infty dt e^{-st} \oint_p dx_{||} \delta(x_{||} - f^t(x_{||})) = T_p \sum_{r=1}^\infty e^{-sT_p r}. \quad (21.16)$$

This integration is a prototype of what needs to be done for each marginal direction, whenever existence of a conserved quantity (energy in Hamiltonian flows, angular momentum, translational invariance, etc.) implies existence of a smooth manifold of equivalent (equivariant) solutions of dynamical equations.

21.2.2 Stability in the transverse directions

Think of the $\tau = 0$ point in above integrals along the cycle p as a choice of a particular Poincaré section. As we have shown in sect. 5.5, the transverse Floquet multipliers do not depend on the choice of a Poincaré section, so ignoring the dependence on $x_{||}(\tau)$ in evaluating the transverse integral in (21.12) is justified. For the transverse integration variables the Jacobian matrix is defined in a reduced Poincaré section $\mathcal{P}$ of fixed $x_{||}$. Linearization of the periodic flow transverse to the orbit yields

$$\int_{\mathcal{P}} dx_{\perp} \delta(x_{\perp} - f_{\perp}^{rT_p}(x)) = \frac{1}{\left| \det(\mathbf{1} - M_p^r) \right|}, \quad (21.17)$$

where M_p is the p -cycle $[d-1 \times d-1]$ transverse monodromy matrix. As in (21.5) we have to assume hyperbolicity, i.e., that the magnitudes of all transverse eigenvalues are bounded away from unity.

Substitution (21.16), (21.17) in (21.12) leads to an expression for $\text{tr } \mathcal{L}^t$ as a sum over all prime cycles p and their repetitions

$$\int_{\epsilon}^{\infty} dt e^{-st} \text{tr } \mathcal{L}^t = \sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r(\beta A_p - s T_p)}}{|\det(\mathbf{1} - M_p^r)|}. \quad (21.18)$$

The $\epsilon \rightarrow 0$ limit of the two expressions for the resolvent, (21.11) and (21.18), now yields the *deterministic trace formula for flows*

$$\sum_{\alpha=0}^{\infty} \frac{1}{s - s_{\alpha}} = \sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r(\beta A_p - s T_p)}}{|\det(\mathbf{1} - M_p^r)|}. \quad (21.19)$$

(If you are worried about the convergence of the resolvent sum, keep the ϵ regularization.)

exercise 21.1

This formula is still another example of the duality between the (local) cycles and (global) eigenvalues. If T_p takes only integer values, we can replace $e^{-s} \rightarrow z$ throughout, so the trace formula for maps (21.9) is a special case of the trace formula for flows. The relation between the continuous and discrete time cases can be summarized as follows:

$$\begin{aligned} T_p &\leftrightarrow n_p \\ e^{-s} &\leftrightarrow z \\ e^{t\mathcal{A}} &\leftrightarrow \mathcal{L}^n. \end{aligned} \quad (21.20)$$

The beauty of trace formulas is that they are coordinate independent: the $|\det(\mathbf{1} - M_p)| = |\det(\mathbf{1} - M^{T_p}(x))|$ and $e^{\beta A_p} = e^{\beta A(x, T_p)}$ contributions to the cycle weight t_p are both independent of the starting periodic point $x \in \mathcal{M}_p$. For the Jacobian matrix M_p this follows from the chain rule for derivatives, and for $e^{\beta A_p}$ from the fact that the integral over $e^{\beta A(x, t)}$ is evaluated along a closed loop. In addition, as we have shown in sect. 5.3, $|\det(\mathbf{1} - M_p)|$ is invariant under smooth coordinate transformations.

We could now proceed to estimate the location of the leading singularity of $\text{tr}(s - \mathcal{A})^{-1}$ by extrapolating finite cycle period truncations of (21.19) by methods such as Padé approximants. However, it pays to first perform a simple resummation which converts this divergence of a trace into a *zero* of a spectral determinant. We shall do this in sect. 22.2, but first a brief refresher of how all this relates to the formula for escape rate (1.8) offered in the introduction might help digest the material.



fast track:

sect. 22, p. 433

21.3 An asymptotic trace formula



In order to illuminate the manipulations of sect. 21.1 and relate them to something we already possess intuition about, we now rederive the heuristic sum of sect. 1.5.1 from the exact trace formula (21.9). The Laplace transforms (21.9) or (21.19) are designed to capture the time $\rightarrow \infty$ asymptotic behavior of the trace sums. By the hyperbolicity assumption (21.5), for $t = T_p r$ large the cycle weight approaches

$$\left| \det(\mathbf{1} - M_p^r) \right| \rightarrow |\Lambda_p|^r, \quad (21.21)$$

where Λ_p is the product of the expanding eigenvalues of M_p . Denote the corresponding approximation to the n th trace (21.6) by

$$\Gamma_n = \sum_i^{(n)} \frac{1}{|\Lambda_i|}, \quad (21.22)$$

and denote the approximate trace formula obtained by replacing the cycle weights $\left| \det(\mathbf{1} - M_p^r) \right|$ by $|\Lambda_p|^r$ in (21.9) by $\Gamma(z)$. Equivalently, think of this as a replacement of the evolution operator (20.14) by a transfer operator (as in example 21.2). For concreteness consider a dynamical system whose symbolic dynamics is complete binary, for example the 3-disk system figure 1.6. In this case distinct periodic points that contribute to the n th periodic points sum (21.7) are labeled by all admissible itineraries composed of sequences of letters $s_i \in \{0, 1\}$:

$$\begin{aligned} \Gamma(z) &= \sum_{n=1}^{\infty} z^n \Gamma_n = \sum_{n=1}^{\infty} z^n \sum_{x_i \in \text{Fix } f^n} \frac{e^{\beta A(x_i, n)}}{|\Lambda_i|} \\ &= z \left\{ \frac{e^{\beta A_0}}{|\Lambda_0|} + \frac{e^{\beta A_1}}{|\Lambda_1|} \right\} + z^2 \left\{ \frac{e^{2\beta A_0}}{|\Lambda_0|^2} + \frac{e^{\beta A_{01}}}{|\Lambda_{01}|} + \frac{e^{\beta A_{10}}}{|\Lambda_{10}|} + \frac{e^{2\beta A_1}}{|\Lambda_1|^2} \right\} \\ &\quad + z^3 \left\{ \frac{e^{3\beta A_0}}{|\Lambda_0|^3} + \frac{e^{\beta A_{001}}}{|\Lambda_{001}|} + \frac{e^{\beta A_{010}}}{|\Lambda_{010}|} + \frac{e^{\beta A_{100}}}{|\Lambda_{100}|} + \dots \right\} \end{aligned} \quad (21.23)$$

Both the cycle averages A_i and the stabilities Λ_i are the same for all points $x_i \in \mathcal{M}_p$ in a cycle p . Summing over repeats of all prime cycles we obtain

$$\Gamma(z) = \sum_p \frac{n_p t_p}{1 - t_p}, \quad t_p = z^{n_p} e^{\beta A_p} / |\Lambda_p|. \quad (21.24)$$

This is precisely our initial heuristic estimate (1.9) (compare also with (18.8), (21.9) and (21.19)). Note that we could not perform such sum over r in the exact trace formula (21.9) as $\left| \det(\mathbf{1} - M_p^r) \right| \neq \left| \det(\mathbf{1} - M_p) \right|^r$; the correct way to resum the exact trace formulas is to first expand the factors $1/|1 - \Lambda_{p,i}|$, as we shall do in (22.8).

section 22.2

If the weights $e^{\beta A(x, n)}$ are multiplicative along the flow, and the flow is hyperbolic, for given β the magnitude of each $|e^{\beta A(x_i, n)} / \Lambda_i|$ term is bounded by some

constant M^n . The total number of cycles grows as 2^n (or as e^{hn} , h = topological entropy, in general), and the sum is convergent for z sufficiently small, $|z| < 1/2M$. For large n the n th level sum (21.6) tends to the leading $\mathcal{L}^n$ eigenvalue e^{ns_0} . Summing this asymptotic estimate level by level

$$\Gamma(z) \approx \sum_{n=1}^{\infty} (ze^{s_0})^n = \frac{ze^{s_0}}{1 - ze^{s_0}} \quad (21.25)$$

we see that we should be able to determine s_0 by determining the smallest value of $z = e^{-s_0}$ for which the cycle expansion (21.24) diverges.

If one is interested only in the leading eigenvalue of $\mathcal{L}$, it suffices to consider the approximate trace $\Gamma(z)$. We will use this fact in sect. 22.3 to motivate the introduction of dynamical zeta functions (22.10), and in sect. 22.5 we shall give the exact relation between the exact and the approximate trace formulas.

Résumé

The description of a chaotic dynamical system in terms of cycles can be visualized as a tessellation of the dynamical system, figure 21.1, with a smooth flow approximated by its *periodic orbit skeleton*, each region M_i centered on a periodic point x_i of the topological period n , and the size of the region determined by the linearization of the flow around the periodic point. The integral over such topologically partitioned state space yields the *deterministic trace formula*

$$\sum_{\alpha=0}^{\infty} \frac{1}{s - s_{\alpha}} = \sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r(\beta A_p - s T_p)}}{|\det(\mathbf{1} - M_p^r)|}.$$

Now that we have a trace formula, one might ask: what is it good for? As it stands, it is a scary divergent formula which relates the unspeakable infinity of global eigenvalues to the unthinkable infinity of local unstable cycles. However, it is a good stepping stone on the way to construction of spectral determinants (to which we turn next), and a first hint that when the going is good, the theory might turn out to be convergent beyond our wildest dreams (chapter 28). In order to implement such formulas, we will have to determine “all” prime cycles. The first step is topological: enumeration of all admissible cycles undertaken in chapter 15. The more onerous enterprise of actually computing the cycles we first approach traditionally, as a numerical task in chapter 16, and then more boldly as a part and parcel of variational foundations of deterministic and quantum dynamics in chapter 34.

Commentary

Remark 21.1. Who’s dunne it? Continuous time flow traces weighted by cycle periods were introduced by Bowen [4] who treated them as Poincaré section suspensions weighted by the “time ceiling” function (3.7). They were used by Parry and Pollicott [7].

Remark 21.2. Flat and sharp traces. In the above formal derivation of trace formulas we cared very little whether our sums were well posed. In the Fredholm theory traces like (21.10) require compact operators with continuous function kernels. This is not the case for our Dirac delta evolution operators: nevertheless, there is a large class of dynamical systems for which our results may be shown to be perfectly legal. In the mathematical literature expressions like (21.6) are called *flat* traces (see the Baladi-Ruelle review [3] and chapter 28). Other names for traces appear as well: for instance, in the context of 1-dimensional mappings, *sharp* traces refer to generalizations of (21.6) where contributions of periodic points are weighted by the Lefschetz sign ± 1 , reflecting whether the periodic point sits on a branch of n th iterate of the map which crosses the diagonal starting from below or starting from above [6]. Such traces are connected to the theory of kneading invariants (see ref. [3] and references therein). Traces weighted by ± 1 sign of the derivative of the fixed point have been used to study the period doubling repeller, leading to high precision estimates of the Feigenbaum constant δ , refs. [1, 2, 5, 8].

References

- [1] R. Artuso, E. Aurell, and P. Cvitanović, “Recycling of strange sets: I. Cycle expansions”, *Nonlinearity* **3**, 325–359 (1990).
- [2] R. Artuso, E. Aurell, and P. Cvitanović, “Recycling of strange sets: II. Applications”, *Nonlinearity* **3**, 361–386 (1990).
- [3] V. Baladi and D. Ruelle, “An extension of the theorem of Milnor and Thurston on the zeta functions of interval maps”, *Ergod. Theor. Dynam. Syst.* **14**, 621–632 (1994).
- [4] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms* (Springer, New York, 1975).
- [5] F. Christiansen, P. Cvitanović, and H. H. Rugh, “The spectrum of the period-doubling operator in terms of cycles”, *J. Phys. A* **23**, L713S–L717S (1990).
- [6] D. Fried, “Lefschetz formula for flows”, in *The Lefschetz Centennial Conference*, Vol. 58, edited by S. Gitler (Amer. Math. Soc., 1987), pp. 19–69.
- [7] W. Parry and M. Pollicott, *Zeta Functions and the Periodic Orbit Structure of Hyperbolic Dynamics*, Vol. 187–188 (Astérisque, Soc. Math. France, 1990).
- [8] M. Pollicott, “A note on the Artuso-Aurell-Cvitanović approach to the Feigenbaum tangent operator”, *J. Stat. Phys.* **62**, 257–267 (1991).

21.4 Examples

Example 21.1. Elliptic stability. Elliptic stability, i.e., a pair of purely imaginary exponents $\Lambda_m = e^{\pm i\theta}$ is excluded by the hyperbolicity assumption. While the contribution of a single repeat of a cycle

$$\frac{1}{(1 - e^{i\theta})(1 - e^{-i\theta})} = \frac{1}{2(1 - \cos \theta)} \quad (21.26)$$

does not make (19.9) diverge, if $\Lambda_m = e^{i2\pi p/r}$ is r th root of unity, $1/\left|\det(\mathbf{1} - M_p')\right|$ diverges. For a generic θ repeats $\cos(r\theta)$ behave badly and by ergodicity $1 - \cos(r\theta)$ is arbitrarily small, $1 - \cos(r\theta) < \epsilon$, infinitely often. This goes by the name of “small divisor problem,” and requires a separate treatment.

[return to p. 422](#)

Example 21.2. A trace formula for transfer operators: For a piecewise-linear map (19.38), we can explicitly evaluate the trace formula. By the piecewise linearity and the chain rule $\Lambda_p = \Lambda_0^{n_0} \Lambda_1^{n_1}$, where the cycle p contains n_0 symbols 0 and n_1 symbols 1. The trace sum is the sum over all periodic points, corresponding to all possible strings of ‘0’s and ‘1’s of length n . The number of strings with $n_0 = m$ ‘0’s and $n_1 = n - n_0$ ‘1’s is given by the usual combinatorial formula. As they all have the same weight $|1 - \Lambda_0^m \Lambda_1^{n-m}|^{-1}$, the trace (21.6) takes the binomial sum form

$$\text{tr } \mathcal{L}^n = \sum_{m=0}^n \binom{n}{m} \frac{1}{|1 - \Lambda_0^m \Lambda_1^{n-m}|}.$$

Now expand the weights in Taylor series

$$\begin{aligned} \text{tr } \mathcal{L}^n &= \sum_{m=0}^n \binom{n}{m} \frac{1}{|\Lambda_0|^m} \frac{1}{|\Lambda_1|^{n-m}} \frac{1}{1 - \Lambda_0^{-m} \Lambda_1^{-n+m}} \\ &= \sum_{k=0}^{\infty} \sum_{m=0}^n \binom{n}{m} \frac{1}{|\Lambda_0|^m \Lambda_0^{km}} \frac{1}{|\Lambda_1|^{n-m} \Lambda_1^{k(n-m)}}. \end{aligned}$$

By the binomial theorem for $(a+b)^n$, for this piecewise-linear map, the trace (21.6) reduces to

$$\text{tr } \mathcal{L}^n = \sum_{k=0}^{\infty} \left(\frac{1}{|\Lambda_0| \Lambda_0^k} + \frac{1}{|\Lambda_1| \Lambda_1^k} \right)^n, \quad (21.27)$$

with eigenvalues

$$e^{s_k} = \frac{1}{|\Lambda_0| \Lambda_0^k} + \frac{1}{|\Lambda_1| \Lambda_1^k}. \quad (21.28)$$

As the simplest example of spectrum for such dynamical system, consider the symmetric piecewise-linear 2-branch repeller (19.38) for which $\Lambda = \Lambda_1 = -\Lambda_0$. In this case all odd eigenvalues vanish, and the even eigenvalues are given by $e^{s_k} = 2/\Lambda^{k+1}$, k even.

[exercise 19.7](#)

Asymptotically the spectrum (21.28) is dominated by the lesser of the two fixed point slopes $\Lambda = \Lambda_0$ (if $|\Lambda_0| < |\Lambda_1|$, otherwise $\Lambda = \Lambda_1$), and the eigenvalues e^{s_k} fall off exponentially as $1/\Lambda^k$, dominated by the single less unstable fixed-point.

[example 28.1](#)

For $k = 0$ this is in agreement with the explicit transfer matrix (19.40) eigenvalues (20.31). The alert reader should experience anxiety at this point. Is it not true that we have already written down explicitly the transfer operator in (19.40), and that it is clear by inspection that it has only *one* eigenvalue $e^{s_0} = 1/|\Lambda_0| + 1/|\Lambda_1|$? The example at hand is one of the simplest illustrations of necessity of defining the space that the operator acts on in order to define the spectrum. The transfer operator (19.40) is the correct operator on the space of functions piecewise constant on the state space partition $\{\mathcal{M}_0, \mathcal{M}_1\}$; on this space the operator indeed has only the eigenvalue e^{s_0} . As we shall see in example 28.1, the full spectrum (21.28) corresponds to the action of the transfer operator on the space of real analytic functions.

The Perron-Frobenius operator trace formula for the piecewise-linear map (19.38) follows from (21.8)

$$\mathrm{tr} \frac{z\mathcal{L}}{1 - z\mathcal{L}} = \frac{z \left(\frac{1}{|\Lambda_0-1|} + \frac{1}{|\Lambda_1-1|} \right)}{1 - z \left(\frac{1}{|\Lambda_0-1|} + \frac{1}{|\Lambda_1-1|} \right)}, \quad (21.29)$$

verifying the trace formula (21.9).

[return to p. 423](#)

Exercises

21.1. $t \rightarrow 0_+$ regularization of eigenvalue sums.



In taking the Laplace transform (21.19) we have ignored the $t \rightarrow 0_+$ divergence, as we do not know how to regularize the delta function kernel in this limit. In the quantum (or heat kernel) case this limit gives rise to the Weyl or Thomas-Fermi mean eigenvalue spacing. Regularize the divergent sum in (21.19) and assign to such volume term some interesting role in the theory of deterministic resonance spectra. E-mail the solution to the authors.

21.2. Deterministic trace formula for flows.



Verify (or improve) the steps in the derivation of the continuous time trace formula

$$\sum_{\alpha=0}^{\infty} \frac{1}{s - s_{\alpha}} = \sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r(\beta A_p - s T_p)}}{\left| \det(\mathbf{1} - M_p^r) \right|}. \quad (21.30)$$

Chapter 22

Spectral determinants

“It seems very pretty,” she said when she had finished it, “but it’s rather hard to understand!” (You see she didn’t like to confess, even to herself, that she couldn’t make it out at all.) “Somehow it seems to fill my head with ideas — only I don’t exactly know what they are!”

—Lewis Carroll, *Through the Looking Glass*

THE PROBLEM with the trace formulas (21.9), (21.19) and (21.24) is that they diverge at $z = e^{-s_0}$, respectively $s = s_0$, i.e., precisely where one would like to use them. While this does not prevent numerical estimation of some “thermodynamic” averages for iterated mappings, in the case of the Gutzwiller trace formula this leads to a perplexing observation that crude estimates of the radius of convergence seem to put the entire physical spectrum out of reach. We shall now cure this problem by thinking, at no extra computational cost; while traces and determinants are formally equivalent, determinants are the tool of choice when it comes to computing spectra. Determinants tend to have larger analyticity domains because if $\text{tr } \mathcal{L}/(1 - z\mathcal{L}) = -\frac{d}{dz} \ln \det(1 - z\mathcal{L})$ diverges at a particular value of z , then $\det(1 - z\mathcal{L})$ might have an isolated zero there, and a zero of a function is easier to determine numerically than its poles.

chapter 28

22.1 Spectral determinants for maps

The eigenvalues z_k of a linear operator are given by the zeros of the determinant

$$\det(1 - z\mathcal{L}) = \prod_k (1 - z/z_k). \quad (22.1)$$

For finite matrices this is the characteristic determinant; for operators this is the Hadamard representation of the *spectral determinant* (sparing the reader from pondering possible regularization factors). Consider first the case of maps, for

which the evolution operator advances the densities by integer steps in time. In this case we can use the formal matrix identity

exercise 4.1

$$\ln \det(1 - M) = \operatorname{tr} \ln(1 - M) = - \sum_{n=1}^{\infty} \frac{1}{n} \operatorname{tr} M^n, \quad (22.2)$$

to relate the spectral determinant of an evolution operator for a map to its traces (21.7), and hence to periodic orbits:

$$\begin{aligned} \det(1 - z\mathcal{L}) &= \exp \left(- \sum_n \frac{z^n}{n} \operatorname{tr} \mathcal{L}^n \right) \\ &= \exp \left(- \sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{z^{n_p r} e^{r\beta A_p}}{|\det(\mathbf{1} - M_p^r)|} \right). \end{aligned} \quad (22.3)$$

Going the other way, the trace formula (21.9) can be recovered from the spectral determinant by taking a derivative

$$\operatorname{tr} \frac{z\mathcal{L}}{1 - z\mathcal{L}} = -z \frac{d}{dz} \ln \det(1 - z\mathcal{L}). \quad (22.4)$$



example 22.1
p. 444

22.2 Spectral determinant for flows

... an analogue of the [Artin-Mazur] zeta function for diffeomorphisms seems quite remote for flows. However we will mention a wild idea in this direction. [...] define $l(\gamma)$ to be the minimal period of γ [...] then define formally (another zeta function!) $Z(s)$ to be the infinite product

$$Z(s) = \prod_{\gamma \in \Gamma} \prod_{k=0}^{\infty} (1 - [\exp l(\gamma)]^{-s-k}).$$

—Stephen Smale, *Differentiable Dynamical Systems*

We write the formula for the spectral determinant for flows by analogy to (22.3)



$$\det(s - \mathcal{A}) = \exp \left(- \sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{r(\beta A_p - s T_p)}}{|\det(\mathbf{1} - M_p^r)|} \right), \quad (22.5)$$

and then check that the trace formula (21.19) is the logarithmic derivative of the



spectral determinant

$$\mathrm{tr} \frac{1}{s - \mathcal{A}} = \frac{d}{ds} \ln \det (s - \mathcal{A}). \quad (22.6)$$



With z set to $z = e^{-s}$ as in (21.20), the spectral determinant (22.5) has the same form for both maps and flows. We refer to (22.5) as *spectral determinant*, as the spectrum of the operator $\mathcal{A}$ is given by the zeros of

$$\det (s - \mathcal{A}) = 0. \quad (22.7)$$

We now note that the r sum in (22.5) is close in form to the expansion of a logarithm. This observation enables us to recast the spectral determinant into an infinite product over periodic orbits as follows:

Let M_p be the p -cycle $[d \times d]$ transverse Jacobian matrix, with eigenvalues $\Lambda_{p,1}, \Lambda_{p,2}, \dots, \Lambda_{p,d}$. Expanding the expanding eigenvalue factors $1/(1 - 1/\Lambda_{p,e})$ and the contracting eigenvalue factors $1/(1 - \Lambda_{p,c})$ in (21.4) as geometric series, substituting back into (22.5), and resumming the logarithms, we find that the spectral determinant is formally given by the infinite product

$$\det (s - \mathcal{A}) = \prod_{k_1=0}^{\infty} \cdots \prod_{l_c=0}^{\infty} \frac{1}{\zeta_{k_1 \cdots l_c}} \quad (22.8)$$

$$1/\zeta_{k_1 \cdots l_c} = \prod_p \left(1 - t_p \frac{\Lambda_{p,e+1}^{l_1} \Lambda_{p,e+2}^{l_2} \cdots \Lambda_{p,d}^{l_c}}{\Lambda_{p,1}^{k_1} \Lambda_{p,2}^{k_2} \cdots \Lambda_{p,e}^{k_e}} \right)$$

$$t_p = t_p(z, s, \beta) = \frac{1}{|\Lambda_p|} e^{\beta A_p - s T_p} z^{n_p}. \quad (22.9)$$

In such formulas t_p is a weight associated with the p cycle (letter t refers to the “local trace” evaluated along the p cycle trajectory), and the index p runs through all distinct prime cycles. Why the factor z^{n_p} ? It is associated with the trace formula (21.9) for maps, whereas the factor $e^{-s T_p}$ is specific to the continuous time trace formula (21.19); according to (21.20) we should use either one or the other. But we have learned in sect. 3.1 that flows can be represented either by their continuous-time trajectories, or by their topological time Poincaré section return maps. In cases when we have good control over the topology of the flow, it is often convenient to insert the z^{n_p} factor into cycle weights, as a formal parameter which keeps track of the topological cycle periods. These factors will assist us in expanding zeta functions and determinants, eventually we shall set $z = 1$. The subscripts e, c indicate that there are e expanding eigenvalues, and c contracting eigenvalues. The observable whose average we wish to compute contributes through the $A(x, t)$ term in the p cycle multiplicative weight $e^{\beta A_p}$. By its definition (20.1), the weight for maps is a product along the periodic points

chapter 23

$$e^{A_p} = \prod_{j=0}^{n_p-1} e^{a(f^j(x_p))},$$

and the weight for flows is an exponential of the integral (20.4) along the cycle

$$e^{A_p} = \exp \left(\int_0^{T_p} a(x(\tau)) d\tau \right).$$

This formula is correct for scalar weighting functions; more general matrix valued weights require a time-ordering prescription as in the Jacobian matrix of sect. 4.1.



example 22.2
p. 444



example 22.3
p. 444

Now we are finally poised to deal with the problem posed at the beginning of chapter 21; how do we actually evaluate the averages introduced in sect. 20.1? The eigenvalues of the dynamical averaging evolution operator are given by the values of s for which the spectral determinant (22.5) of the evolution operator (20.14) vanishes. If we can compute the leading eigenvalue $s_0(\beta)$ and its derivatives, we are done. Unfortunately, the infinite product formula (22.8) is no more than a shorthand notation for the periodic orbit weights contributing to the spectral determinant; more work will be needed to bring such formulas into a tractable form. This shall be accomplished in chapter 23, but here it is natural to introduce still another variant of a determinant, the dynamical zeta function.

22.3 Dynamical zeta functions

It follows from sect. 21.1.1 that if one is interested only in the leading eigenvalue of $\mathcal{L}^t$, the size of the p cycle neighborhood can be approximated by $1/|\Lambda_p|^r$, the dominant term in the $rT_p = t \rightarrow \infty$ limit, where $\Lambda_p = \prod_e \Lambda_{p,e}$ is the product of the expanding eigenvalues of the Jacobian matrix M_p . With this replacement the spectral determinant (22.5) is replaced by the *dynamical zeta function*



$$1/\zeta = \exp \left(- \sum_p \sum_{r=1}^{\infty} \frac{1}{r} t_p^r \right) \quad (22.10)$$

that we have already derived heuristically in sect. 1.5.2. Resumming the logarithms using $-\sum_r t_p^r/r = \ln(1 - t_p)$ we obtain the *Euler product representation* of the dynamical zeta function:

$$1/\zeta = \prod_p (1 - t_p). \quad (22.11)$$

In order to simplify the notation, we usually omit the explicit dependence of $1/\zeta$, t_p on z, s, β whenever the dependence is clear from the context.



The approximate trace formula (21.24) plays the same role *vis-à-vis* the dynamical zeta function (22.6)

$$\Gamma(s) = \frac{d}{ds} \ln \zeta^{-1} = \sum_p \frac{T_p t_p}{1 - t_p}, \quad (22.12)$$

as the exact trace formula (21.19) plays *vis-à-vis* the spectral determinant (22.5). The heuristically derived dynamical zeta function of sect. 1.5.2 now re-emerges as the $1/\zeta_{0\ldots 0}(z)$ part of the *exact* spectral determinant; other factors in the infinite product (22.8) affect the non-leading eigenvalues of $\mathcal{L}$.

In summary, the dynamical zeta function (22.11) associated with the flow $f^t(x)$ is defined as the product over all prime cycles p . The quantities, T_p , n_p and Λ_p , denote the period, topological length and product of the expanding Floquet multipliers of prime cycle p , A_p is the Birkhoff sum $a(x)$ evaluated on a single traversal of cycle p (see (20.4)), s is a variable dual to the time t , z is dual to the discrete “topological” time n , and $t_p(z, s, \beta)$ denotes the local trace over the cycle p . We have included the factor z^{n_p} in the definition of the cycle weight in order to keep track of the number of times a cycle traverses the surface of section. The dynamical zeta function is useful because the term

$$1/\zeta(s) = 0 \quad (22.13)$$

when $s = s_0$, Here s_0 is the leading eigenvalue of $\mathcal{L}^t = e^{t\mathcal{A}}$, which is often all that is necessary for application of this equation. The above argument completes our derivation of the trace and determinant formulas for deterministic chaotic flows. In chapters that follow we shall make these formulas tangible by working out a series of simple examples.

The remainder of this chapter offers examples of zeta functions.



fast track:
chapter 23, p. 448

22.4 False zeros

Compare (A22.9) with the Euler product (22.11). For simplicity consider two equal scales, $|\Lambda_0| = |\Lambda_1| = e^\lambda$. Our task is to determine the leading zero $z = e^\gamma$ of the Euler product. It is a novice error to assume that the infinite Euler product (22.11) vanishes whenever one of its factors vanishes. If that were true, each factor $(1 - z^{n_p}/|\Lambda_p|)$ would yield

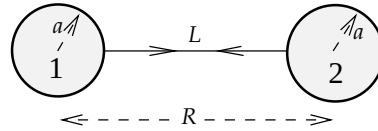
$$0 = 1 - e^{n_p(\gamma - \lambda_p)}, \quad (22.14)$$

so the escape rate γ would equal the Floquet exponent of a unstable cycle, one eigenvalue $\gamma = \gamma_p$ for each prime cycle p . This is false! The exponentially growing number of cycles with growing period conspires to shift the zeros of the infinite product. The correct formula follows from (A22.9)

$$0 = 1 - e^{\gamma - \lambda + h}, \quad h = \ln 2. \quad (22.15)$$

This particular formula for the escape rate is a special case of a general relation between escape rates, Lyapunov exponents and entropies that is not yet included

Figure 22.1: A game of pinball consisting of two disks of equal size in a plane, radius a , center-to-center separation R , with its only periodic orbit of period $T = 2L$. (A. Wirzba)



into ChaosBook. Physically this means that the escape induced by the repulsion by each unstable fixed point is diminished by the rate of backscatter from other repelling regions, i.e., the entropy h ; the positive entropy of orbits shifts the “false zeros” $z = e^{\lambda_p}$ of the Euler product (22.11) to the true zero $z = e^{\lambda-h}$.

22.5 Spectral determinants vs. dynamical zeta functions

In sect. 22.3 we derived the dynamical zeta function as an approximation to the spectral determinant. Here we relate dynamical zeta functions to spectral determinants *exactly*, by showing that a dynamical zeta function can be expressed as a ratio of products of spectral determinants.

The elementary identity for d -dimensional matrices

$$1 = \frac{1}{\det(1 - M)} \sum_{k=0}^d (-1)^k \text{tr}(\wedge^k M), \quad (22.16)$$

inserted into the exponential representation (22.10) of the dynamical zeta function, relates the dynamical zeta function to *weighted* spectral determinants.



example 22.4
p. 445



example 22.5
p. 445



example 22.6
p. 445

22.6 All too many eigenvalues?

What does the 2-dimensional hyperbolic Hamiltonian flow spectral determinant (22.23) tell us? Consider one of the simplest conceivable hyperbolic flows: the game of pinball of figure 22.1 consisting of two disks of equal size in a plane. There is only one periodic orbit, with the period T and expanding eigenvalue Λ given by elementary considerations (see exercise 16.6), and the resonances $\det(s_\alpha - \mathcal{A}) = 0$, $\alpha = \{k, n\}$ plotted in figure 22.2:

$$s_\alpha = -(k+1)\lambda + n\frac{2\pi i}{T}, \quad n \in \mathbb{Z}, k \in \mathbb{Z}_+, \quad \text{multiplicity } k+1, \quad (22.17)$$

can be read off the spectral determinant (22.23) for a single unstable cycle:

$$\det(s - \mathcal{A}) = \prod_{k=0}^{\infty} \left(1 - e^{-sT}/|\Lambda|^k\right)^{k+1}. \quad (22.18)$$

In the above $\lambda = \ln|\Lambda|/T$ is the cycle Lyapunov exponent. For an open system, the real part of the eigenvalue s_α gives the decay rate of α th eigenstate, and the



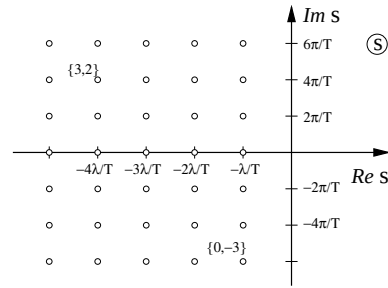


Figure 22.2: The deterministic resonances $\alpha = \{k, n\}$ (22.17) for a 2-disk game of pinball.

imaginary part gives the “node number” of the eigenstate. The negative real part of s_α indicates that the resonance is unstable, and the decay rate in this simple case (zero entropy) equals the cycle Lyapunov exponent.

Rapidly decaying eigenstates with large negative $\text{Re } s_\alpha$ are not a problem, but as there are eigenvalues arbitrarily far in the imaginary direction, this might seem like all too many eigenvalues. However, they are necessary - we can check this by explicit computation of the right hand side of (21.19), the trace formula for flows:

$$\begin{aligned}
 \sum_{\alpha=0}^{\infty} e^{s_\alpha t} &= \sum_{k=0}^{\infty} \sum_{n=-\infty}^{\infty} (k+1) e^{-(k+1)\lambda t + i2\pi n t/T} \\
 &= \sum_{k=0}^{\infty} (k+1) \left(\frac{1}{|\Lambda| \Lambda^k} \right)^{t/T} \sum_{n=-\infty}^{\infty} e^{i2\pi n t/T} = \sum_{k=0}^{\infty} \frac{k+1}{|\Lambda|^r \Lambda^{kr}} \sum_{r=-\infty}^{\infty} \delta(r - t/T) \\
 &= T \sum_{r=-\infty}^{\infty} \frac{\delta(t - rT)}{|\Lambda|^r (1 - 1/\Lambda^r)^2}. \tag{22.19}
 \end{aligned}$$

Hence, the two sides of the trace formula (21.19) are verified. The formula is fine for $t > 0$; for $t \rightarrow 0_+$, however, sides are divergent and need regularization.

The reason why such sums do not occur for maps is that for discrete time we work with the variable $z = e^s$, so an infinite strip along $\text{Im } s$ maps into an annulus in the complex z plane, and the Dirac delta sum in the above is replaced by the Kronecker delta sum in (21.7). In the case at hand there is only one time scale T , and we could just as well replace s by the variable $z = e^{-sT}$. In general, a continuous time flow has an infinity of irrationally related cycle periods, and the resonance arrays are more irregular, cf. figure 23.1.



in depth:

appendix A22, p. 1007

Résumé

The eigenvalues of evolution operators are given by the zeros of corresponding determinants, and one way to evaluate determinants is to expand them in terms

of traces, using the matrix identity $\log \det = \text{tr} \log$. Traces of evolution operators can be evaluated as integrals over Dirac delta functions, and in this way the spectra of evolution operators are related to periodic orbits. The spectral problem is now recast into a problem of determining zeros of either the *spectral determinant*

$$\det(s - \mathcal{A}) = \exp \left(- \sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{(\beta A_p - s T_p)r}}{|\det(\mathbf{1} - M'_p)|} \right),$$

or the leading zeros of the *dynamical zeta function*

$$1/\zeta = \prod_p (1 - t_p), \quad t_p = \frac{1}{|\Lambda_p|} e^{\beta A_p - s T_p}.$$

The spectral determinant is the tool of choice in actual calculations, as it has superior convergence properties (this will be discussed in chapter 28 and is illustrated, for example, by table 23.2). In practice both spectral determinants and dynamical zeta functions are preferable to trace formulas because they yield the eigenvalues more readily; the main difference is that while a trace diverges at an eigenvalue and requires extrapolation methods, determinants vanish at s corresponding to an eigenvalue s_α , and are analytic in s in an open neighborhood of s_α .

The critical step in the derivation of the periodic orbit formulas for spectral determinants and dynamical zeta functions is the hyperbolicity assumption (21.5) that no cycle stability eigenvalue is marginal, $|\Lambda_{p,i}| \neq 1$. By dropping the prefactors in (1.5), we have given up on any possibility of recovering the precise distribution of the initial x (return to the past is rendered moot by the chaotic mixing and the exponential growth of errors), but in exchange we gain an effective description of the asymptotic behavior of the system. The pleasant surprise (to be demonstrated in chapter 23) is that the infinite time behavior of an unstable system turns out to be as easy to determine as its short time behavior.



Commentary

Remark 22.1. Piecewise monotone maps. A partial list of cases for which the transfer operator is well defined: the expanding Hölder case, weighted subshifts of finite type, expanding differentiable case, see Bowen [6]; expanding holomorphic case, see Ruelle [15]; piecewise monotone maps of the interval, see Hofbauer and Keller [10] and Baladi and Keller [4].

Remark 22.2. Smale's wild idea. Smale's wild idea quoted on page 434 was technically wrong because 1) the Selberg zeta function yields the spectrum of a quantum mechanical Laplacian rather than the deterministic resonances, 2) the spectral determinant weights are different from what Smale conjectured, as the individual cycle weights also depend on the stability of the cycle, 3) the formula is not dimensionally correct, as k is an

integer and s represents inverse time. Only for spaces of constant negative curvature do all cycles have the same Lyapunov exponent $\lambda = \ln |\Lambda_p|/T_p$. In this case, one can normalize time so that $\lambda = 1$, and the factors e^{-sT_p}/Λ_p^k in (22.8) simplify to $s^{-(s+k)T_p}$, as intuited in Smale's quote on page 434 (where $l(\gamma)$ is the cycle period). Nevertheless, Smale's intuition was remarkably on the target.

Remark 22.3. Is this a generalization of the Fourier analysis? Fourier analysis is a theory of the space $\leftrightarrow$ eigenfunction duality for dynamics on a circle. The way in which periodic orbit theory generalizes Fourier analysis to nonlinear flows is discussed in ref. [13], a very readable introduction to the Selberg Zeta function.

Remark 22.4. Zeta functions, antecedents. For a function to be deserving of appellation “zeta,” one expects it to have an Euler product representation (22.11), and perhaps also satisfy a functional equation. Various kinds of zeta functions are reviewed in refs. [11, 12, 25]. Historical antecedents of the dynamical zeta function are the fixed-point counting functions introduced by Weil [26], Lefschetz [8] and Artin and Mazur [1], and the determinants of transfer operators of statistical mechanics [21]. In his review article Smale [22] already intuited, by analogy to the Selberg Zeta function, that the spectral determinant is the right generalization for continuous time flows. Ruelle [17] discusses various guises of the dynamical zeta functions, Sharp [19] offers a comprehensive survey of periodic orbits of hyperbolic flows, and Baladi [2] is currently the most in-depth monograph on the subject.

In dynamical systems theory, dynamical zeta functions arise naturally only for piecewise linear mappings; for smooth flows the natural object for the study of classical and quantal spectra are the spectral determinants. Ruelle derived the relation (22.3) between spectral determinants and dynamical zeta functions, but since he was motivated by the Artin-Mazur zeta function (18.17) and the statistical mechanics analogy, he did not consider the spectral determinant to be a more natural object than the dynamical zeta function. This has been put right in papers on “flat traces” [3, 5].

The nomenclature has not settled down yet; what we call evolution operators here is elsewhere called transfer operators [18], Perron-Frobenius operators [7, 14, 24] and/or Ruelle-Araki operators. Here we refer to kernels such as (20.14) as evolution operators. We follow Ruelle in usage of the term “dynamical zeta function,” but elsewhere in the literature function (22.11) may be called the Ruelle, Artin-Mazur-Ruelle, or Smale-Ruelle zeta function. Ruelle [16] points out that the corresponding transfer operator T was never considered by either Perron or Frobenius; a more appropriate designation would be the Ruelle-Araki operator. Determinants similar to or identical with our spectral determinants are sometimes called Selberg Zetas, Selberg-Smale zetas [9] (Gaspard [9] credits Smale [23] - but why, dynamical spectral determinant was not derived by either Selberg or Smale? and we cannot find any other zeta for the flows but the “wild” one quoted on page 434), functional determinants, Fredholm determinants, or even - to maximize confusion - dynamical zeta functions [20]. A Fredholm determinant is a notion that applies only to trace class operators - as we consider here a somewhat wider class of operators, we prefer to refer to their determinants loosely as “spectral determinants.”

appendix A40.3

References

- [1] M. Artin and B. Mazur, “On periodic points”, *Ann. Math.* **81**, 82–99 (1965).
- [2] V. Baladi, *Dynamical Zeta Functions and Dynamical Determinants for Hyperbolic Maps, A Functional Approach* (Springer, Berlin, 2017).
- [3] V. Baladi, D. R. A. Kitaev, and S. Semmes, “Sharp determinants and kneading operators for holomorphic maps”, *Proc. Steklov Inst. of Math.* **216**, 186–228 (1997).
- [4] V. Baladi and G. Keller, “Zeta functions and transfer operators for piecewise monotone transformations”, *Commun. Math. Phys.* **127**, 459–477 (1990).
- [5] V. Baladi and D. Ruelle, “Sharp determinants”, *Inv. Math.* **123**, 553–574 (1996).
- [6] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms* (Springer, New York, 1975).
- [7] S.-J. Chang and J. Wright, “Transitions and distribution functions for chaotic systems”, *Phys. Rev. A* **23**, 1419–1433 (1981).
- [8] D. Fried, “Lefschetz formula for flows”, in *The Lefschetz Centennial Conference*, Vol. 58, edited by S. Gitler (Amer. Math. Soc., 1987), pp. 19–69.
- [9] P. Gaspard, *Chaos, Scattering and Statistical Mechanics* (Cambridge Univ. Press, Cambridge UK, 1998).
- [10] F. Hofbauer and G. Keller, “Ergodic properties of invariant measures for piecewise monotonic transformations”, *Math. Z.* **180**, 119–140 (1982).
- [11] N. E. Hurt, “Zeta functions and periodic orbit theory: A review”, *Results in Mathematics* **23**, 55–120 (1993).
- [12] K. Itô, *Encyclopedic Dictionary of Mathematics* (MIT press, 1993).
- [13] H. P. McKean, “Selberg’s trace formula as applied to a compact Riemann surface”, *Commun. Pure Appl. Math.* **25**, 225–246 (1972).
- [14] Y. Oono and Y. Takahashi, “Chaos, external noise and Fredholm theory”, *Progr. Theor. Phys.* **63**, 1804–1807 (1980).
- [15] D. Ruelle, “Zeta-functions for expanding maps and Anosov flows”, *Inv. Math.* **34**, 231–242 (1976).
- [16] D. Ruelle, *Functional determinants related to dynamical systems and the thermodynamic formalism*, tech. rep. IHES-P-95-30 (Inst. Hautes Etud. Sci., Bures-sur-Yvette, 1995).
- [17] D. Ruelle, “Dynamical zeta functions and transfer operators”, *Notices Amer. Math. Soc.* **95**, 887–895 (2002).
- [18] D. Ruelle, *Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics*, 2nd ed. (Cambridge Univ. Press, Cambridge UK, 2004).

- [19] R. Sharp, “Periodic orbits of hyperbolic flows”, in *On Some Aspects of the Theory of Anosov Systems*, edited by G. A. Margulis (Springer, Berlin, 2004), pp. 73–138.
- [20] M. Sieber and F. Steiner, “Classical and quantum mechanics of a strongly chaotic billiard system”, *Physica D* **44**, 248–266 (1990).
- [21] Y. G. Sinai, “Gibbs measures in ergodic theory”, *Russ. Math. Surv.* **27**, 21–69 (1972).
- [22] S. Smale, “Differentiable dynamical systems”, *Bull. Amer. Math. Soc.* **73**, 747–817 (1967).
- [23] S. Smale, *The Mathematics of Time* (Springer, New York, 1980).
- [24] Y. Takahashi and Y. Oono, “Towards the statistical mechanics of chaos”, *Progr. Theor. Phys* **71**, 851–854 (1984).
- [25] A. Voros, “Zeta functions in geometry”, *Adv. Studies in Pure Math., Math. Soc. Japan* **21**, 327–358 (1992).
- [26] A. Weil, “Numbers of solutions of equations in finite fields”, *Bull. Amer. Math. Soc* **55**, 497–508 (1949).

22.7 Examples

Example 22.1. Spectral determinants of transfer operators:



For a piecewise-linear map (19.38) with a finite Markov partition, an explicit formula for the spectral determinant follows by substituting the trace formula (21.27) into (22.3):

$$\det(1 - z\mathcal{L}) = \prod_{k=0}^{\infty} \left(1 - \frac{t_0}{\Lambda_0^k} - \frac{t_1}{\Lambda_1^k} \right), \quad (22.20)$$

where $t_s = z/|\Lambda_s|$. The eigenvalues are necessarily the same as in (21.28), which we already determined from the trace formula (21.9).

The exponential spacing of eigenvalues guarantees that the spectral determinant (22.20) is an entire function. It is this property that generalizes to piecewise smooth flows with finite Markov partitions, and singles out spectral determinants rather than the trace formulas or dynamical zeta functions as the tool of choice for evaluation of spectra.

[return to p. 434](#)

Example 22.2. Expanding 1-dimensional map:



For expanding 1-dimensional mappings the spectral determinant (22.8) takes the form

$$\det(1 - z\mathcal{L}) = \prod_p \prod_{k=0}^{\infty} \left(1 - t_p/\Lambda_p^k \right), \quad t_p = \frac{e^{\beta A_p}}{|\Lambda_p|} z^{n_p}. \quad (22.21)$$

[return to p. 436](#)

Example 22.3. Two-degrees of freedom Hamiltonian flows: For a 2-degrees of freedom Hamiltonian flow the energy conservation eliminates one phase-space variable, and restriction to a Poincaré section eliminates the marginal longitudinal eigenvalue $\Lambda = 1$, so a periodic orbit of 2-degrees of freedom hyperbolic Hamiltonian flow (or of a 1-degree of freedom hyperbolic Hamiltonian map) has one expanding transverse eigenvalue Λ , $|\Lambda| > 1$, and one contracting transverse eigenvalue $1/\Lambda$. The weight in (21.4) is expanded as follows:

$$\frac{1}{|\det(\mathbf{1} - M_p^r)|} = \frac{1}{|\Lambda|^r (1 - 1/\Lambda_p^r)^2} = \frac{1}{|\Lambda|^r} \sum_{k=0}^{\infty} \frac{k+1}{\Lambda_p^{kr}}. \quad (22.22)$$

The spectral determinant exponent can be resummed,

$$-\sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{(\beta A_p - s T_p)r}}{|\det(\mathbf{1} - M_p^r)|} = \sum_{k=0}^{\infty} (k+1) \log \left(1 - \frac{e^{\beta A_p - s T_p}}{|\Lambda_p| \Lambda_p^k} \right),$$

and the spectral determinant for a 2-dimensional hyperbolic Hamiltonian flow rewritten as an infinite product over prime cycles

$$\det(s - \mathcal{A}) = \prod_p \prod_{k=0}^{\infty} \left(1 - t_p/\Lambda_p^k \right)^{k+1}. \quad (22.23)$$

[exercise 28.4](#)
[return to p. 436](#)

Example 22.4. Dynamical zeta function in terms of determinants, 1-dimensional maps: For 1-dimensional maps the identity

$$1 = \frac{1}{(1 - 1/\Lambda)} - \frac{1}{\Lambda} \frac{1}{(1 - 1/\Lambda)}$$

substituted into (22.10) yields an expression for the dynamical zeta function for 1-dimensional maps as a ratio of two spectral determinants

$$1/\zeta = \frac{\det(1 - z\mathcal{L})}{\det(1 - z\mathcal{L}_{(1)})} \quad (22.24)$$

where the cycle weight in $\mathcal{L}_{(1)}$ is given by replacement $t_p \rightarrow t_p/\Lambda_p$. As we shall see in chapter 28, this establishes that for nice hyperbolic flows $1/\zeta$ is meromorphic, with poles given by the zeros of $\det(1 - z\mathcal{L}_{(1)})$. The dynamical zeta function and the spectral determinant have the same zeros, although in exceptional circumstances some zeros of $\det(1 - z\mathcal{L}_{(1)})$ might be cancelled by coincident zeros of $\det(1 - z\mathcal{L})$. Hence even though we have derived the dynamical zeta function in sect. 22.3 as an “approximation” to the spectral determinant, the two contain the same spectral information.

[return to p. 438](#)

Example 22.5. Dynamical zeta function in terms of determinants, 2-dimensional Hamiltonian maps: For 2-dimensional Hamiltonian flows the above identity yields

$$\frac{1}{|\Lambda|} = \frac{1}{|\Lambda|(1 - 1/\Lambda)^2} (1 - 2/\Lambda + 1/\Lambda^2),$$

so

$$1/\zeta = \frac{\det(1 - z\mathcal{L}) \det(1 - z\mathcal{L}_{(2)})}{\det(1 - z\mathcal{L}_{(1)})^2}. \quad (22.25)$$

This establishes that for nice 2-dimensional hyperbolic flows the dynamical zeta function is meromorphic.

[return to p. 438](#)

Example 22.6. Dynamical zeta functions for 2-dimensional Hamiltonian flows: The relation (22.25) is not particularly useful for our purposes. Instead we insert the identity

$$1 = \frac{1}{(1 - 1/\Lambda)^2} - \frac{2}{\Lambda} \frac{1}{(1 - 1/\Lambda)^2} + \frac{1}{\Lambda^2} \frac{1}{(1 - 1/\Lambda)^2}$$

into the exponential representation (22.10) of $1/\zeta_k$, and obtain

$$1/\zeta_k = \frac{\det(1 - z\mathcal{L}_{(k)}) \det(1 - z\mathcal{L}_{(k+2)})}{\det(1 - z\mathcal{L}_{(k+1)})^2}. \quad (22.26)$$

Even though we have no guarantee that $\det(1 - z\mathcal{L}_{(k)})$ are entire, we do know that the upper bound on the leading zeros of $\det(1 - z\mathcal{L}_{(k+1)})$ lies strictly below the leading zeros of $\det(1 - z\mathcal{L}_{(k)})$, and therefore we expect that for 2-dimensional Hamiltonian flows the dynamical zeta function $1/\zeta_k$ generically has a *double* leading pole coinciding with the leading zero of the $\det(1 - z\mathcal{L}_{(k+1)})$ spectral determinant. This might fail if the poles and leading eigenvalues come in wrong order, but we have not encountered such situations in our numerical investigations. This result can also be stated as follows: the theorem establishes that the spectral determinant (22.23) is entire, and also implies that the poles in $1/\zeta_k$ must have the right multiplicities to cancel in the $\det(1 - z\mathcal{L}) = \prod 1/\zeta_k^{k+1}$ product.

[return to p. 438](#)

Exercises

22.1. **Spectrum of the ‘golden mean’ pruned map.** (medium - exercise 18.7 continued)

- (a) Determine an expression for $\text{tr } \mathcal{L}^n$, the trace of powers of the Perron-Frobenius operator (19.10) acting on the space of real analytic functions for the tent map of exercise 18.7.
- (b) Show that the spectral determinant for the Perron-Frobenius operator is

$$\det(1 - z\mathcal{L}) = \prod_{k \text{ even}} \left(1 - \frac{z}{\Lambda^{k+1}} - \frac{z^2}{\Lambda^{2k+2}}\right) \times \prod_{k \text{ odd}} \left(1 + \frac{z}{\Lambda^{k+1}} + \frac{z^2}{\Lambda^{2k+2}}\right). \quad (22.27)$$

22.2. **Dynamical zeta functions from transition graphs.** Extend sect. 18.3 to evaluation of dynamical zeta functions for piecewise linear maps with finite transition graphs. This generalizes the results of exercise A22.3.

22.3. **Zeros of infinite products.** Determination of the quantities of interest by periodic orbits involves working with infinite product formulas.

- (a) Consider the infinite product

$$F(z) = \prod_{k=0}^{\infty} (1 + f_k(z))$$

where the functions f_k are “sufficiently nice.” This infinite product can be converted into an infinite sum by the use of a logarithm. Use the properties of infinite sums to develop a sensible definition of infinite products.

- (b) If z^* is a root of the function F , show that the infinite product diverges when evaluated at z^* .
- (c) How does one compute a root of a function represented as an infinite product?
- (d) Let p be all prime cycles of the binary alphabet $\{0, 1\}$. Apply your definition of $F(z)$ to the infinite product

$$F(z) = \prod_p \left(1 - \frac{z^{n_p}}{\Lambda^{n_p}}\right)$$

- (e) Are the roots of the factors in the above product the zeros of $F(z)$?

(Per Rosenqvist)

22.4. **Dynamical zeta functions** as ratios of spectral determinants. (medium) Show that the zeta function

$$1/\zeta(z) = \exp \left(- \sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{z^{n_p}}{|\Lambda_p|^r} \right)$$

can be written as the ratio

$$1/\zeta(z) = \det(1 - z\mathcal{L}_{(0)})/\det(1 - z\mathcal{L}_{(1)}),$$

where $\det(1 - z\mathcal{L}_{(s)}) = \prod_p \prod_{k=0}^{\infty} (1 - z^{n_p}/|\Lambda_p|\Lambda_p^{k+s})$.

- 22.5. **Dynamical zeta function for maps.** In this problem we will compare the dynamical zeta function and the spectral determinant. Compute the exact dynamical zeta function for the skew full tent map (19.44)

$$1/\zeta(z) = \prod_{p \in P} \left(1 - \frac{z^{n_p}}{|\Lambda_p|}\right).$$

What are its roots? Do they agree with those computed in exercise 19.7?

- 22.6. **Dynamical zeta functions for Hamiltonian maps.** Starting from

$$1/\zeta(s) = \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{1}{r} t_p^r\right)$$

for a 2-dimensional Hamiltonian map. Using the equality

$$1 = \frac{1}{(1 - 1/\Lambda)^2} (1 - 2/\Lambda + 1/\Lambda^2),$$

show that

$$1/\zeta = \det(1 - \mathcal{L}) \det(1 - \mathcal{L}_{(2)})/\det(1 - \mathcal{L}_{(1)})^2.$$

In this expression $\det(1 - z\mathcal{L}_{(k)})$ is the expansion one gets by replacing $t_p \rightarrow t_p/\Lambda_p^k$ in the spectral determinant.

- 22.7. **Riemann ζ function.** The Riemann ζ function is defined as the sum

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \quad s \in \mathbb{C}.$$

- (a) Use factorization into primes to derive the Euler product representation

$$\zeta(s) = \prod_p \frac{1}{1 - p^{-s}}.$$

The dynamical zeta function exercise 22.11 is called a “zeta” function because it shares the form of the Euler product representation with the Riemann zeta function.

- (b) (Not trivial:) For which complex values of s is the Riemann zeta sum convergent?
(c) Are the zeros of the terms in the product, $s = -\ln p$, also the zeros of the Riemann ζ function? If not, why not?

- 22.8. **Finite truncations.** (easy) Suppose we have a 1-dimensional system with complete binary dynamics, where the stability of each orbit is given by a simple multiplicative rule:

$$\Lambda_p = \Lambda_0^{n_{p,0}} \Lambda_1^{n_{p,1}}, \quad n_{p,0} = \#0 \text{ in } p, \quad n_{p,1} = \#1 \text{ in } p,$$

so that, for example, $\Lambda_{00101} = \Lambda_0^3 \Lambda_1^2$.

- (a) Compute the dynamical zeta function for this system; perhaps by creating a transfer matrix analogous to (19.40), with the right weights.
(b) Compute the finite p truncations of the cycle expansion, i.e. take the product only over the p up to given period with $n_p \leq N$, and expand as a series in z

$$\prod_p \left(1 - \frac{z^{n_p}}{|\Lambda_p|}\right).$$

Do they agree? If not, how does the disagreement depend on the truncation period N ?

Chapter 23

Cycle expansions

Recycle... It's the Law!

—Poster, New York City Department of Sanitation

WHEN WE SET OUT on this journey, we had promised to teach you something profound that *your professor does not know*. Well, this chapter is the chapter. If your professor knows *cycle formulas for dynamical averages*, please send us her name, and we'll feature it in ChaosBook. They look like cumulants, but when you start to take them apart you realize how brilliant they are - your professor would not guess their form even if he wrote 1000 and 7 *Physical Review Letters* about it. Takes 20 some chapters of hard study to start to understand them, and who has time for that?



The Euler product representations of spectral determinants (22.8) and dynamical zeta functions (22.11) are really only a shorthand notation - the zeros of the individual factors are *not* the zeros of the zeta function, and the convergence of these objects is far from obvious. Now we shall give meaning to dynamical zeta functions and spectral determinants by expanding them as *cycle expansions*, which are series representations ordered by increasing topological cycle period, with products in (22.8), (22.11) expanded as sums over *pseudo-cycles*, products of weights t_p of contributing cycles. The zeros of correctly truncated cycle expansions yield the desired leading eigenvalues of evolution operators, and the expectation values of observables are given by the cycle averaging formulas obtained from the partial derivatives of dynamical zeta functions (or spectral determinants).

For reasons of pedagogy in what follows everything is first explained in terms of dynamical zeta functions: they aid us in developing 'shadowing' intuition about the geometrical meaning of cycle expansions. For actual calculations, we recommend the spectral determinant cycle expansions of sects. 23.2.2 and 23.5.2. While the shadowing is less transparent, and the weights calculation is an iterative numerical algorithm, these expansions use full analytic information about the flow, and can have better convergence properties than the dynamical zeta functions. For example, as we shall show in chapter 28, even when a spectral determinant (22.5)

is entire and calculations are super-exponentially convergent, cycle expansion of the corresponding dynamical zeta function (22.24) has a finite radius of convergence and captures only the leading eigenvalue, at exponentially convergent rate.

23.1 Pseudo-cycles and shadowing

How are periodic orbit formulas such as (22.11) evaluated? We start by computing the periods and Floquet multipliers of the shortest cycles. This always requires numerical work, such as searches for periodic solutions via Newton's method; we shall assume for the purpose of this discussion that the numerics is under control, and that *all* short cycles up to a given (topological) period have been found. Examples of the data required for application of periodic orbit formulas are the lists of cycles given in exercise 7.2 and table 34.1. Sadly, it is not enough to set a computer to blindly troll for invariant solutions, and blithely feed those into the formulas that will be given here. The reason that this chapter is numbered 23 and not 6, is that understanding the geometry of the non-wandering set is a prerequisite to good estimation of dynamical averages: one has to identify cycles that belong to a given ergodic component (whose symbolic dynamics and shadowing is organized by its transition graph), and discard the isolated cycles and equilibria that do not take part in the asymptotic dynamics. It is important not to miss any short cycles, as the calculation is as accurate as the shortest cycle dropped - including cycles longer than the shortest omitted does not improve the accuracy (more precisely, the calculation improves, but so little as not to be worth while).

chapter 16

Given a set of periodic orbits, we can compute their weights t_p and expand the dynamical zeta function (22.11) as a formal power series,

$$1/\zeta = \prod_p (1 - t_p) = 1 - \sum'_{\{p_1 p_2 \dots p_k\}} (-1)^{k+1} t_{p_1} t_{p_2} \dots t_{p_k} \quad (23.1)$$

where the prime on the sum indicates that the sum is over all distinct non-repeating combinations of prime cycles.

As we shall frequently use such sums, let us denote by $t_\pi = (-1)^{k+1} t_{p_1} t_{p_2} \dots t_{p_k}$ an element of the set of all distinct products of the prime cycle weights t_p , and label each such *pseudo-cycle* by

$$\pi = p_1 + p_2 + \dots + p_k \quad (23.2)$$

The formal power series (23.1) is now compactly written as

$$1/\zeta = 1 - \sum'_{\pi} t_\pi. \quad (23.3)$$

For $k > 1$, the signed products t_π are weights of *pseudo-cycles*; they are sequences of shorter cycles that shadow a cycle with the symbol sequence $p_1 p_2 \dots p_k$ along the segments $p_1, p_2, \dots, p_k$, as in figure 1.11. The symbol $\sum'$ denotes the restricted sum, for which any given prime cycle p contributes at most once to a given pseudo-cycle weight t_π .

The pseudo-cycle weight, i.e., the product of weights (22.9) of prime cycles comprising the pseudo-cycle,

$$t_\pi = (-1)^{k+1} \frac{1}{|\Lambda_\pi|} e^{\beta A_\pi - s T_\pi} z^{n_\pi}, \quad (23.4)$$

depends on the pseudo-cycle Birkhoff sum A_π , the period T_π , the stability Λ_π ,

remark 5.1

$$\begin{aligned} \Lambda_\pi &= \Lambda_{p_1} \Lambda_{p_2} \cdots \Lambda_{p_k}, & T_\pi &= T_{p_1} + \dots + T_{p_k} \\ A_\pi &= A_{p_1} + \dots + A_{p_k}, & n_\pi &= n_{p_1} + \dots + n_{p_k}, \end{aligned} \quad (23.5)$$

and, when available, the topological period n_π .

23.1.1 Curvature expansions

The simplest example is the pseudo-cycle sum for a system described by a complete binary symbolic dynamics. In this case the Euler product (22.11) is given by

$$\begin{aligned} 1/\zeta &= (1 - t_0)(1 - t_1)(1 - t_{01})(1 - t_{001})(1 - t_{011}) \\ &\quad \times (1 - t_{0001})(1 - t_{0011})(1 - t_{0111})(1 - t_{00001})(1 - t_{00011}) \\ &\quad \times (1 - t_{00101})(1 - t_{00111})(1 - t_{01011})(1 - t_{01111}) \dots \end{aligned} \quad (23.6)$$

(see table 18.1), and the first few terms of the expansion (23.3) ordered by increasing total pseudo-cycle length are:

$$\begin{aligned} 1/\zeta &= 1 - t_0 - t_1 - t_{01} - t_{001} - t_{011} - t_{0001} - t_{0011} - t_{0111} - \dots \\ &\quad + t_{0+1} + t_{0+01} + t_{01+1} + t_{0+001} + t_{0+011} + t_{001+1} + t_{011+1} \\ &\quad - t_{0+01+1} - \dots \end{aligned} \quad (23.7)$$

We refer to such series representation of a dynamical zeta function or a spectral determinant, expanded as a sum over pseudo-cycles, and ordered by increasing cycle period and instability, as a *cycle expansion*.

The next step is the key step: regroup the terms into the dominant *fundamental* contributions t_f and the decreasing *curvature* corrections $\hat{c}_n$, each $\hat{c}_n$ split into prime cycles p of period $n_p=n$ grouped together with pseudo-cycles whose full itineraries build up the itinerary of p . For the binary case this regrouping is given by

$$\begin{aligned} 1/\zeta &= 1 - t_0 - t_1 - [(t_{01} - t_{0+1})] - [(t_{001} - t_{0+01}) + (t_{011} - t_{01+1})] \\ &\quad - [(t_{0001} - t_{0+001}) + (t_{0111} - t_{011+1}) \\ &\quad \quad + (t_{0011} - t_{001+1} - t_{0+011} + t_{0+01+1})] - \dots \\ &= 1 - \sum_f t_f - \sum_n \hat{c}_n. \end{aligned} \quad (23.8)$$

All terms in this expansion up to period $n_p = 6$ are given in table 23.1. We refer to such regrouped series as *curvature expansions*, because the shadowed combinations $[\dots]$ vanish identically for piecewise-linear maps with nice partitions, such as the ‘full tent map’ of figure 19.5.

This separation into ‘fundamental’ and ‘curvature’ parts of cycle expansions is possible *only* for dynamical systems whose symbolic dynamics has finite grammar. The fundamental cycles t_0, t_1 have no shorter approximations; they are the “building blocks” of the dynamics in the sense that all longer orbits can be approximately pieced together from them. The fundamental part of a cycle expansion is given by the sum of the products of all non-intersecting loops of the associated transition graph, discussed in chapter 17. The terms grouped in brackets $[\dots]$ are the curvature corrections; the terms grouped in parentheses $(\dots)$ are combinations of longer cycles and corresponding sequences of “shadowing” pseudo-cycles, as in figure 1.11. If all orbits are weighted equally ($t_p = z^{n_p}$), such combinations cancel exactly, and the dynamical zeta function reduces to the topological polynomial (18.17). If the flow is continuous and smooth, orbits of similar symbolic dynamics will traverse the same neighborhoods and will have similar weights, and the weights in such combinations will almost cancel. The utility of cycle expansions of dynamical zeta functions and spectral determinants, in contrast to naive averages over periodic orbits such as the trace formulas discussed in sect. 27.4, lies precisely in this organization into nearly canceling combinations: cycle expansions are dominated by short cycles, with longer cycles giving exponentially decaying corrections.

section 18.3
section 23.6

More often than not, good symbolic dynamics for a given flow is either not available, or its grammar is not finite, or the convergence of cycle expansions is affected by nonhyperbolic regions of state space. In those cases truncations such as the *stability cutoff* of sect. 23.7 and sect. 29.3.4 might be helpful. The idea is to truncate the cycle expansion by including only the pseudo-cycles such that $|\Lambda_{p_1} \cdots \Lambda_{p_k}| \leq \Lambda_{\max}$, with the cutoff $\Lambda_{\max}$ equal to or greater than the most unstable Λ_p in the data set.

In what follows, we shall introduce two cycle averaging formulas, one based on dynamical zeta functions and the other on spectral determinants. (Frequently used, but inferior ‘level sums’ shall be discussed in sect. 27.4.)

23.2 Construction of cycle expansions

Due to the lack of factorization of the determinant in the denominator of the full pseudo-cycle weight in (21.19),

$$\det(\mathbf{1} - M_{p_1 p_2}) \neq \det(\mathbf{1} - M_{p_1}) \det(\mathbf{1} - M_{p_2}),$$

the cycle expansions for the spectral determinant (22.8) are somewhat less transparent than is the case for the dynamical zeta functions, so we postpone their evaluation to sect. 23.2.2. Sect. 23.2.1 is a pedagogical warmup. In actual calculations, implementing the spectral determinant cycle expansions of sect. 23.2.2 is

Table 23.1: The binary curvature expansion (23.8) up to period 6, listed in such a way that the sum of terms along the p th horizontal line is the curvature $\hat{c}_p$ associated with a prime cycle p , or a combination of prime cycles such as the $t_{100101} + t_{100110}$ pair.

- t_0			
- t_1			
- t_{10}	+ $t_1 t_0$		
- t_{100}	+ t_{10+0}		
- t_{101}	+ t_{10+1}		
- t_{1000}	+ t_{100+0}		
- t_{1001}	+ t_{100+1}	+ t_{110+0}	- t_{1+10+0}
- t_{1011}	+ t_{101+1}		
- t_{10000}	+ t_{1000+0}		
- t_{10001}	+ t_{1001+0}	+ t_{1000+1}	- $t_{0+100+1}$
- t_{10010}	+ t_{100+10}		
- t_{10101}	+ t_{101+10}		
- t_{10011}	+ t_{1011+0}	+ t_{1001+1}	- $t_{0+101+1}$
- t_{10111}	+ t_{1011+1}		
- t_{100000}	+ $t_{10000+0}$		
- t_{100001}	+ $t_{10001+0}$	+ $t_{10000+1}$	- $t_{0+1000+1}$
- t_{100010}	+ $t_{10010+0}$	+ $t_{1000+10}$	- $t_{0+100+10}$
- t_{100011}	+ $t_{10011+0}$	+ $t_{10001+1}$	- $t_{0+1001+1}$
- t_{100101}	- t_{100110}	+ $t_{10010+1}$	+ $t_{10110+0}$
	+ $t_{10+1001}$	+ $t_{100+101}$	- $t_{0+10+101} - t_{1+10+100}$
- t_{101110}	+ $t_{10110+1}$	+ $t_{1011+10}$	- $t_{1+101+10}$
- t_{100111}	+ $t_{10011+1}$	+ $t_{10111+0}$	- $t_{0+1011+1}$
- t_{101111}	+ $t_{10111+1}$		

recommended. Correct objects are spectral determinants, and as using the correct object costs exactly the same as using the approximations, why settle for less?

23.2.1 Evaluation of dynamical zeta functions

Cycle expansions of dynamical zeta functions are evaluated numerically by first computing the weights $t_p = t_p(\beta, s)$ of all prime cycles p of topological period $n_p \leq N$, for given fixed β and s . Denote by the subscript (i) the i th prime cycle computed, ordered by the topological period $n_{(i)} \leq n_{(i+1)}$. The dynamical zeta function $1/\zeta_N$ truncated to $n_p \leq N$ cycles is computed recursively, by multiplying

$$1/\zeta_{(i)} = 1/\zeta_{(i-1)}[1 - t_{(i)}z^{n_{(i)}}], \quad (23.9)$$

and truncating the expansion at each step to a finite polynomial in z^n , $n \leq N$. The result is the N th order polynomial approximation

$$1/\zeta_N = 1 - \sum_{n=1}^N c_n z^n. \quad (23.10)$$

In other words, a cycle expansion is a Taylor expansion in the dummy variable z , where each term in the sum is raised to the topological cycle period. If both the number of cycles and their individual weights grow not faster than exponentially with the cycle period, and we multiply the weight of each cycle p by a factor z^{n_p} , the cycle expansion converges for sufficiently small $|z|$. If the symbolic dynamics grammar is finite, the truncation cutoff N has to be larger than the period of

longest cycle in the transition graph (18.13), for the salubrious effect of shadowing cancelations to kick in. If that is the case, further increases in N yield the exponentially decreasing corrections $\hat{c}_n$ in (23.8).

If the dynamics is given by an iterated mapping, the leading zero of (23.10) as a function of z yields the leading eigenvalue of the appropriate evolution operator. For continuous time flows, z is a dummy variable that we set to $z = 1$, and the leading eigenvalue of the evolution operator is given by the leading zero of $1/\zeta(s, \beta(s))$ as function of s .

23.2.2 Evaluation of traces and spectral determinants

We commence the cycle expansion evaluation of a spectral determinant by computing the trace formula (21.9) or (21.19). The weight of prime cycle p repeated r times is

$$t_p(z, \beta, r) = \frac{e^{r\beta A_p} z^{r n_p}}{|\det(\mathbf{1} - M_p^r)|} \quad (\text{discrete time}) \quad (23.11)$$

$$t_p(s, \beta, r) = \frac{e^{r(\beta A_p - s T_p)}}{|\det(\mathbf{1} - M_p^r)|} \quad (\text{continuous time}). \quad (23.12)$$

For *discrete time*, the trace formula (21.9) truncated to all prime cycles p and their repeats r such that $n_p r \leq N$,

$$\text{tr} \frac{z\mathcal{L}}{1 - z\mathcal{L}} \Big|_N = \sum_{n=1}^N C_n z^n, \quad C_n = \text{tr} \mathcal{L}^n, \quad (23.13)$$

is computed as a polynomial in z by adding a cycle at the time:

$$\text{tr} \frac{z\mathcal{L}}{1 - z\mathcal{L}} \Big|_{(i)} = \text{tr} \frac{z\mathcal{L}}{1 - z\mathcal{L}} \Big|_{(i-1)} + n_{(i)} \sum_{r=1}^{n_{(i)} r \leq N} t_{(i)}(z, \beta, r).$$

For *continuous time*, we assume that the method of Poincaré sections assigns each cycle a topological period n_p . Then the trace formula (21.19) is also organized as a polynomial

$$\text{tr} \frac{1}{s - \mathcal{A}} \Big|_N = \sum_{n=1}^N C_n z^n, \quad (23.14)$$

computed as:

$$\text{tr} \frac{1}{s - \mathcal{A}} \Big|_{(i)} = \text{tr} \frac{1}{s - \mathcal{A}} \Big|_{(i-1)} + T_{(i)} \sum_{r=1}^{n_{(i)} r \leq N} t_{(i)}(s, \beta, r) z^{n_p r}$$

The periodic orbit data set (23.5) consists of the list of the cycle periods T_p , the cycle Floquet multipliers $\Lambda_{p,1}, \Lambda_{p,2}, \dots, \Lambda_{p,d}$, and the cycle averages of the observable A_p for all prime cycles p such that $n_p \leq N$. The coefficient of $z^{n_p r}$ is then

evaluated numerically for the given parameter values (β, s) . Always compute the leading eigenvalue of the evolution operator first, i.e., the escape rate $\gamma = -s_0$, in order to use it in calculation of averages of sect. 23.5 as a weight $e^{\gamma T(i)}$ in (23.13). Now that we have an expansion for the trace formula (21.8) as a power series, we compute the N th order approximation to the spectral determinant (22.3),

$$\det(1 - z\mathcal{L})|_N = 1 - \sum_{n=1}^N Q_n z^n, \quad Q_n = n\text{th cumulant}, \quad (23.15)$$

(compare with (18.10)) as follows. The logarithmic derivative relation (22.4) yields section 41.2.1

$$\left(\text{tr} \frac{z\mathcal{L}}{1 - z\mathcal{L}} \right) \det(1 - z\mathcal{L}) = -z \frac{d}{dz} \det(1 - z\mathcal{L})$$

$$(C_1 z + C_2 z^2 + \dots)(1 - Q_1 z - Q_2 z^2 - \dots) = Q_1 z + 2Q_2 z^2 + 3Q_3 z^3 \dots$$

so the n th order term of the spectral determinant cycle (or in this case, the cumulant) expansion is given recursively by the convolution trace formula expansion coefficients

$$Q_n = \frac{1}{n} (C_n - C_{n-1}Q_1 - \dots - C_1 Q_{n-1}), \quad Q_1 = C_1. \quad (23.16)$$

appendix A20.1

Given the trace formula (23.13) truncated to z^N , we now also have the spectral determinant truncated to z^N .

The same code can also be reused to compute the dynamical zeta function cycle expansion (23.10), by replacing $\prod (1 - \Lambda_{(i),j}^r)$ in (23.13) by the product of expanding eigenvalues $\Lambda_{(i)} = \prod_e \Lambda_{(i),e}$. section 22.3

A few points concerning different cycle averaging formulas:

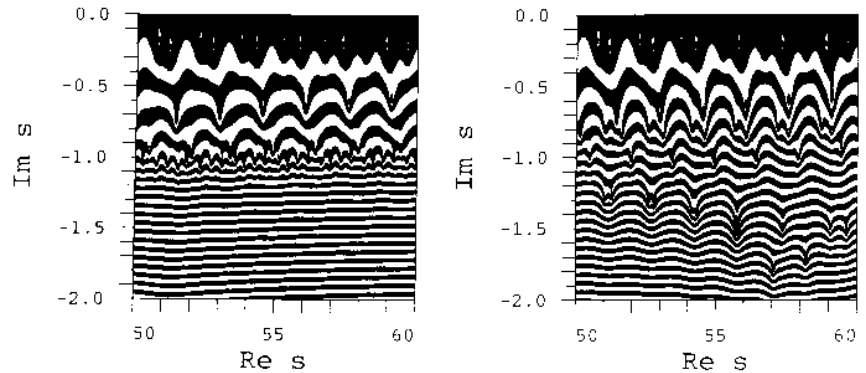
- The dynamical zeta functions is an approximation to spectral determinant that yields only the leading eigenvalue of the evolution operator. The cycle weights depend only on the product of expanding $|\Lambda_i|$ Floquet multipliers, so signs do not matter. For hyperbolic flows they converge exponentially with increasing cycle periods.
- spectral determinants weights in (22.3) contain $1/|1 - \Lambda_i|$ factors, so for them signs of Floquet multipliers Λ_i do matter. With finite grammar the leading eigenvalue converges super-exponentially in cycle period.

Note that while the dynamical zeta functions weights use only the expanding Floquet multipliers $|\Lambda_e|$, for spectral determinants the weights are of form $|1 - \Lambda_j|$, both expanding and contracting directions contribute, and the signs of multipliers do matter. That's why ChaosBook everywhere tracks multipliers Λ_j , rather than Floquet exponents λ_j . λ 's belong to equilibria, periodic orbits require multipliers. That's the way cookie crumbles. For very high-dimensional flows (such as unstable periodic solutions of Navier-Stokes equations), usually only a subset of the most unstable / least contracting Floquet multipliers is known. As long as the contracting Floquet

Table 23.2: The 3-disk repeller escape rates computed from cycle expansions of the spectral determinant (22.5) and the dynamical zeta function (22.11), as functions of the maximal cycle period N . The disk-disk center separation to disk radius ratio is $R:a$, and the $\det(s - \mathcal{A})$ is an estimate of the deterministic escape rate computed from the spectral determinant cycle expansion in the fundamental domain. For larger disk-disk separations, the dynamics is more uniform, as illustrated by the faster convergence. Convergence of spectral determinant $\det(s - \mathcal{A})$ is super-exponential, see chapter 28. For comparison, the $1/\zeta(s)$ column lists estimates from the fundamental domain dynamical zeta function cycle expansion (23.8), and the $1/\zeta(s)_{3\text{-disk}}$ column lists estimates from the full 3-disk cycle expansion (25.54). The convergence of the fundamental domain dynamical zeta function is significantly slower than the convergence of the corresponding spectral determinant, and the full (unfactorized) 3-disk dynamical zeta function has still poorer convergence. (P.E. Rosenqvist.)

$R:a$	N	$\det(s - \mathcal{A})$	$1/\zeta(s)$	$1/\zeta(s)_{3\text{-disk}}$
6	1	0.39	0.407	
	2	0.4105	0.41028	0.435
	3	0.410338	0.410336	0.4049
	4	0.4103384074	0.4103383	0.40945
	5	0.4103384077696	0.4103384	0.410367
	6	0.410338407769346482	0.4103383	0.410338
	7	0.4103384077693464892		0.4103396
	8	0.410338407769346489338468		
	9	0.4103384077693464893384613074		
	10	0.4103384077693464893384613078192		
3	1	0.41		
	2	0.72		
	3	0.675		
	4	0.67797		
	5	0.677921		
	6	0.6779227		
	7	0.6779226894		
	8	0.6779226896002		
	9	0.677922689599532		
	10	0.67792268959953606		

Figure 23.1: Example scans in the complex s plane: contour plots of the logarithm of the absolute values of (left) $1/\zeta(s)$, (right) spectral determinant $\det(s - \mathcal{A})$ for the 3-disk system, separation $R : a = 6$. The A_1 subspace is evaluated numerically. The eigenvalues of the evolution operator $\mathcal{L}$ are given by the centers of elliptic neighborhoods of the rapidly narrowing rings. While the dynamical zeta function is analytic on the $\text{Im } s \geq -1$ half-plane, the spectral determinant is entire and reveals further families of zeros. (P.E. Rosenqvist)



multipliers omitted from the weights in (23.13) are sufficiently strongly contracting, the errors introduced by replacement $|1 - \Lambda_j^r| \rightarrow 1$ for such eigenvalues should be negligible.

- The least enlightened are the ‘level sum’ cycle averaging formulas. There is no point in using them, except that they have to be mentioned (here in sect. 27.4), as there is voluminous literature that uses them.
- Other formulas published in physics literature are likely to be wrong.

remark 23.1

If the set of computed periodic orbits is incomplete, and their Floquet multipliers inaccurate, distinctions between different cycle averaging formulas are academic, as there are not sufficiently many cycles to start worrying about what expansion converges faster.

23.3 Periodic orbit averaging

The first cycle expansion calculation should always be the determination of the leading eigenvalue of the evolution operator, calculated as follows. After the prime cycles and the pseudo-cycles have been grouped into subsets of equal topological period, the dummy variable can be set equal to $z = 1$. With $z = 1$, the expansion (23.15) constitutes the cycle expansion (22.5) for the spectral determinant $\det(s - \mathcal{A})$. We vary s in cycle weights, and determine α th eigenvalue s_α (20.18) by finding $s = s_\alpha$ for which (23.15) vanishes. As an example, the convergence of a leading eigenvalue for a nice hyperbolic system is illustrated in table 23.2 by the list of pinball escape rates $\gamma = -s_0$ estimates computed from truncations of (23.8) and (23.15) to different maximal cycle periods.

chapter 28

The pleasant surprise, to be explained in chapter 28, is that one can prove that the coefficients in these cycle expansions decay exponentially or even faster, because of the analyticity of $\det(s - \mathcal{A})$ or $1/\zeta(s)$, for s values well beyond those for which the corresponding trace formula (21.19) diverges.



example 23.1
p. 471

Our next task will be to compute long-time averages of observables. Three situations arise, two of them equal in practice:

- (i) The system is bounded, and we have all cycles up to some cutoff: always start by testing the cycle expansion sum rules of sect. 23.4.
- (ii) The system is unbounded, and averages have to be computed over a repeller whose natural measure is obtained by balancing local instability with the global escape rate $\gamma = -s_0$, as in sect. 20.3.
- (iii) The system is bounded, but we only have a repelling set consisting of a subset of unstable cycles embedded into the bounded strange attractor. Best one can do is to treat this as an open system, case (iii). That assigns a stationary natural measure to neighborhoods of the solutions used, the local instabilities balanced by a weight that includes escape rate $\exp(\gamma T_p)$. Whether use of this measure improves averages as one increases the stability cutoff depends on whether the longer cycles explore qualitatively different regions of state space not visited by the shorter (fundamental) cycles, or only revisit already known regions (curvature corrections).

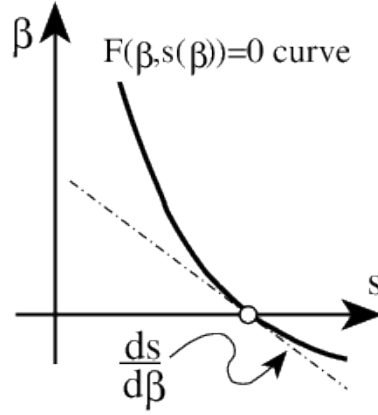
23.4 Flow conservation sum rules

If a dynamical system is bounded, so that all trajectories remain confined for all times, the escape rate (27.8) vanishes $\gamma = -s_0 = 0$, and the leading eigenvalue of the Perron-Frobenius operator (19.10) (evolution operator with $\beta = 0$) is simply $\exp(-t\gamma) = 1$. Conservation of material flow thus implies that for bounded flows cycle expansions of dynamical zeta functions and spectral determinants satisfy exact *flow conservation* sum rules:

$$\begin{aligned} 1/\zeta(0,0) &= 1 + \sum_{\pi}' \frac{(-1)^k}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|} = 0 \\ F(0,0) &= 1 - \sum_{n=1}^{\infty} Q_n(0,0) = 0 \end{aligned} \tag{23.17}$$

obtained by setting $s = 0$ in (23.18), (23.19) with cycle weights $t_p = e^{-sT_p}/|\Lambda_p| \rightarrow 1/|\Lambda_p|$. These sum rules depend neither on the cycle periods T_p nor on the observable $a(x)$ under investigation, but only on the cycle stabilities $\Lambda_{p,1}, \Lambda_{p,2}, \dots, \Lambda_{p,d}$. Their significance is purely geometric; they are a measure of how well periodic orbits tessellate state space, as in figure 1.10. Conservation of material flow provides a first and very useful test of the quality of finite cycle period truncations and is something that you should always check when constructing a cycle expansion for a bounded flow.

Figure 23.2: The eigenvalue condition is satisfied on the curve $F = 0$ on the (β, s) plane. The expectation value of the observable (20.11) is given by the slope of the curve.



23.5 Cycle formulas for dynamical averages

Want to learn some useful safety maneuvers? Or perhaps you'd like to become a more able mechanic? Or have bike safety questions answered? Or eat pizza? Then sign up for Enlightened Cycling!

— Bike GT: Cycling around Georgia Tech

The eigenvalue conditions for the dynamical zeta function (23.3) and the spectral determinant (23.15),

$$0 = 1 - \sum_{\pi}' t_{\pi}, \quad t_{\pi} = t_{\pi}(\beta, s(\beta)) \quad (23.18)$$

$$0 = 1 - \sum_{n=1}^{\infty} Q_n, \quad Q_n = Q_n(\beta, s(\beta)), \quad (23.19)$$

are implicit equations for an eigenvalue $s = s(\beta)$ of the form $0 = F(\beta, s(\beta))$. The eigenvalue $s = s(\beta)$ as a function of β is sketched in figure 23.2; this condition is satisfied on the curve $F = 0$. The cycle averaging formulas for the slope and curvature of $s(\beta)$ are obtained as in (20.11) by taking derivatives of the eigenvalue condition. Evaluated along $F = 0$, by the chain rule the first derivative yields

$$\begin{aligned} 0 &= \frac{d}{d\beta} F(\beta, s(\beta)) \\ &= \frac{\partial F}{\partial \beta} + \frac{ds}{d\beta} \frac{\partial F}{\partial s} \Big|_{s=s(\beta)} \implies \frac{ds}{d\beta} = - \frac{\partial F}{\partial \beta} / \frac{\partial F}{\partial s}, \end{aligned} \quad (23.20)$$

and the second derivative of $F(\beta, s(\beta)) = 0$ yields

$$\frac{d^2 s}{d\beta^2} = - \left[\frac{\partial^2 F}{\partial \beta^2} + 2 \frac{ds}{d\beta} \frac{\partial^2 F}{\partial \beta \partial s} + \left(\frac{ds}{d\beta} \right)^2 \frac{\partial^2 F}{\partial s^2} \right] / \frac{\partial F}{\partial s}. \quad (23.21)$$

Denoting expectations as in (A20.2) by

$$\begin{aligned}\langle A \rangle_F &= - \frac{\partial F}{\partial \beta} \Big|_{\beta, s=s(\beta)}, & \langle T \rangle_F &= \frac{\partial F}{\partial s} \Big|_{\beta, s=s(\beta)}, \\ \langle A^2 \rangle_F &= - \frac{\partial^2 F}{\partial \beta^2} \Big|_{\beta, s=s(\beta)}, & \langle TA \rangle_F &= \frac{\partial^2 F}{\partial s \partial \beta} \Big|_{\beta, s=s(\beta)},\end{aligned}\quad (23.22)$$

the mean cycle expectation value of A , the mean cycle period, and second derivatives of F computed for $F(\beta, s(\beta)) = 0$, we obtain the cycle averaging formulas for the expectation of the observable (20.11) and for its (generalized) diffusion constant (or, more generally, diffusion tensor):

$$\langle a \rangle = \frac{\langle A \rangle_F}{\langle T \rangle_F} \quad (23.23)$$

$$\Delta = \frac{1}{\langle T \rangle_F} \langle (A - T \langle a \rangle)^2 \rangle_F, \quad (23.24)$$

and so forth for higher cumulants. These formulas are the central result of periodic orbit theory. We now show that for each choice of the function $F(\beta, s)$ in (23.3) and (23.15) (also the trace, or ‘level sum’ of (27.15)), the above quantities have explicit cycle expansions.



section A20.1



23.5.1 Dynamical zeta function cycle averaging formulas

For the dynamical zeta function condition (23.18), the cycle averaging formulas (23.20), (23.24) require one to evaluate derivatives of dynamical zeta functions at a given eigenvalue. Substituting the cycle expansion (23.3) for the dynamical zeta function we obtain

$$\begin{aligned}\langle A \rangle_\zeta &:= - \frac{\partial}{\partial \beta} \frac{1}{\zeta} = \sum' A_\pi t_\pi \\ \langle T \rangle_\zeta &:= \frac{\partial}{\partial s} \frac{1}{\zeta} = \sum' T_\pi t_\pi, \quad \langle n \rangle_\zeta := -z \frac{\partial}{\partial z} \frac{1}{\zeta} = \sum' n_\pi t_\pi,\end{aligned}\quad (23.25)$$

where the subscript in $\langle \cdots \rangle_\zeta$ stands for the dynamical zeta function average over prime cycles, A_π , T_π , and n_π given by (23.4) are evaluated on pseudo-cycles (23.5), and pseudo-cycle weights $t_\pi = t_\pi(z, \beta, s(\beta))$ are evaluated at the eigenvalue $s(\beta)$. In most applications $\beta = 0$, and $s(\beta)$ of interest is typically the leading eigenvalue $s_0 = s_0(0)$ of the evolution generator $\mathcal{A}$.

For bounded flows the leading eigenvalue (the escape rate) vanishes, $s(0) = 0$, the exponent $\beta A_\pi - s T_\pi$ in (23.4) vanishes, so the cycle expansions take a simple form

$$\langle A \rangle_\zeta = \sum'_{\pi} (-1)^{k+1} \frac{A_{p_1} + A_{p_2} \cdots + A_{p_k}}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}, \quad (23.26)$$

where analogous formulas hold for $\langle T \rangle_\zeta$, $\langle n \rangle_\zeta$.



example 23.2
p. 471

The cycle averaging formulas for the expectation of observable $\langle a \rangle$ follow by substitution into (23.24). Assuming zero mean $\langle a \rangle = 0$, the cycle expansion (23.15) for the variance $\langle (A - \langle A \rangle)^2 \rangle_\zeta$ is given by

$$\langle A^2 \rangle_\zeta = \sum' (-1)^{k+1} \frac{(A_{p_1} + A_{p_2} \cdots + A_{p_k})^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}. \quad (23.27)$$

23.5.2 Spectral determinant cycle expansions

The dynamical zeta function cycle expansions have a particularly simple structure, with the shadowing apparent already by a term-by-term inspection of table 23.2. For “nice” hyperbolic systems, shadowing ensures exponential convergence of the dynamical zeta function cycle expansions. This, however, is not the best achievable convergence. As will be explained in chapter 28, for nice hyperbolic systems the spectral determinant constructed from the same cycle database is entire, and its cycle expansion converges faster than exponentially. The fastest convergence is attained by the spectral determinant cycle expansion (23.19) and its derivatives. In this case the $\partial/\partial s$, $\partial/\partial \beta$ derivatives are computed recursively, by taking derivatives of the spectral determinant cycle expansion contributions (23.13) and (23.16).

section 28.5

The cycle averaging formulas are exact, and highly convergent for nice hyperbolic dynamical systems. An example of their utility is the cycle expansion formula for the Lyapunov exponent of example 23.3. Further applications of cycle expansions will be discussed in chapter 27.

23.5.3 Continuous vs. discrete mean return time

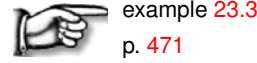
Sometimes it is convenient to compute an expectation value along a flow in continuous time, and sometimes it might be easier to compute it in discrete time, from a return map. Return times (3.1) might vary wildly, and it is not at all clear that the continuous and discrete Birkhoff averages are related in any simple way. As we shall now show, the relationship turns out to be both elegantly simple, and totally general.

exercise 23.12

The mean cycle period $\langle T \rangle_F$ fixes the normalization of the unit of time; it can be interpreted as the average near recurrence or the average first return time. For example, if we have evaluated a billiard expectation value $\langle a \rangle = \langle A \rangle_F / \langle T \rangle_F$ in terms of continuous time, and would like to also have the corresponding average $\langle a \rangle_{\text{dscr}} = \langle A \rangle_F / \langle n \rangle_F$ measured in discrete time, given by the number of reflections off billiard walls, the two averages are related by

$$\langle a \rangle_{\text{dscr}} = \langle a \rangle \langle T \rangle_F / \langle n \rangle_F, \quad (23.28)$$

where $\langle n \rangle_F$ the average of the number of bounces n_p along the cycle p is given by is (23.25).



example 23.3
p. 471

For 2-dimensional Hamiltonian flows such as our game of pinball (see example 22.3), there is only one expanding eigenvalue and (23.44) applies as written. However, in dimensions higher than one, a correct calculation of Lyapunov exponents requires a bit of sophistication.

23.6 Cycle expansions for finite alphabets

A finite transition graph like the one given in figure 17.3 (d) is a compact encoding of the transition matrix for a given subshift. It is a sparse matrix, and the associated determinant (18.33) can be written by inspection: it is the sum of all possible partitions of the graph into products of non-intersecting loops, with each loop carrying a minus sign:

$$\det(1 - T) = 1 - t_0 - t_{0011} - t_{0001} - t_{00011} + t_{0+0011} + t_{0011+0001} \quad (23.29)$$

The simplest application of this determinant is the evaluation of the topological entropy; if we set $t_p = z^{n_p}$, where n_p is the period of the p -cycle, the determinant reduces to the topological polynomial (18.34).

The determinant (23.29) is exact for the finite graph figure 17.3 (e), as well as for the associated finite-dimensional transfer operator of example 20.4. For the associated (infinite dimensional) evolution operator, it is the beginning of the cycle expansion of the corresponding dynamical zeta function:

$$\begin{aligned} 1/\zeta = & 1 - t_0 - t_{0011} - t_{0001} + t_{0001+0011} \\ & -(t_{00011} - t_{0+0011} + \dots \text{curvatures}) \dots \end{aligned} \quad (23.30)$$

The cycles $\bar{0}$, $\overline{0001}$ and $\overline{0011}$ are the *fundamental* cycles introduced in (23.8); they are not shadowed by any combinations of shorter cycles. All other cycles appear together with their shadows (for example, the $t_{00011} - t_{0+0011}$ combination, see figure 1.11) and yield exponentially small corrections for hyperbolic systems. For cycle counting purposes, both t_{ab} and the pseudo-cycle combination $t_{a+b} = t_a t_b$ in (23.3) have the same weight $z^{n_a+n_b}$, so all curvature combinations $t_{ab} - t_{a+b}$ vanish exactly, and the topological polynomial (18.17) offers a quick way of checking the fundamental part of a cycle expansion.

The splitting of cycles into the fundamental cycles and the curvature corrections depends on balancing long cycles t_{ab} against their pseudo-trajectory shadows $t_a t_b$. If the $\overline{ab}$ cycle or either of the shadows $\bar{a}$, $\bar{b}$ do not to exist, such curvature cancelation is unbalanced.

The most important lesson of the pruning of the cycle expansions is that prohibition of a finite subsequence imbalances the head of a cycle expansion and increases the number of the fundamental cycles in (23.8). Hence the pruned expansions are expected to start converging only *after* all fundamental cycles have been incorporated - in the last example, the cycles $\bar{1}$, $\bar{10}$, $\bar{10100}$, $\bar{1011100}$. Without cycle expansions, no such crisp and clear cut definition of the fundamental set of scales is available.

Because topological zeta functions reduce to polynomials for finite grammars, only a few fundamental cycles exist and long cycles can be grouped into curvature combinations. For example, the fundamental cycles in exercise 11.3 are the three 2-cycles that bounce back and forth between two disks and the two 3-cycles that visit every disk. Of all cycles, the 2-cycles have the smallest Floquet exponent, and the 3-cycles the largest. It is only after these fundamental cycles have been included that a cycle expansion is expected to start converging smoothly, i.e., only for n larger than the periods of the fundamental cycles are the curvatures $\hat{c}_n$ (in expansion (23.8)), a measure of the deviations between long orbits and their short cycle approximations, expected to fall off rapidly with n .

23.7 Stability ordering of cycle expansions

There is never a second chance. Most often there is not even the first chance.

—John Wilkins

(C.P. Dettmann and P. Cvitanović)

We have judiciously deployed the 3-disk pinball, with its simple grammar, to motivate the periodic orbit theory. Most dynamical systems of interest, however, have infinite grammar, so at any order in z a cycle expansion may contain unmatched terms that do not fit neatly into the almost canceling curvature corrections. Similarly, for the intermittent systems that we shall discuss in sect. 29.3.4, curvature corrections are not small in general, so again the cycle expansions may converge slowly. For such systems, schemes that collect the pseudocycle terms according to some criterion other than the topology of the flow may converge faster than expansions based on the topological period.

All chaotic systems exhibit some degree of shadowing, and a good truncation criterion should do its best to respect the shadowing as much as possible. If a long cycle is shadowed by two or more shorter cycles and the flow is smooth, the periods and the Floquet exponents will be additive in sense that the period of the longer cycle is approximately the sum of the shorter cycle periods. Similarly, as stability is multiplicative, shadowing is approximately preserved by including all terms with pseudo-cycle stability

$$|\Lambda_{p_1} \cdots \Lambda_{p_k}| \leq \Lambda_{\max} \quad (23.31)$$

and ignoring any pseudo-cycles that are less stable.

Two such schemes for ordering cycle expansions that approximately respect shadowing are truncations by the pseudocycle period and the stability ordering that we shall discuss here. In these schemes, a dynamical zeta function or a spectral determinant is expanded. One keeps all terms for which the period, action or stability for a combination of cycles (pseudo-cycles) is less than a given cutoff.

Settings in which stability ordering may be preferable to ordering by topological cycle period are the cases of bad grammar, of intermittency, and of partial cycle data sets.

23.7.1 Stability ordering for bad grammars

For generic flows it is often not clear what partition of state space generates the “optimal” symbolic dynamics. Stability ordering does not require understanding dynamics in such detail: if you can find the cycles, you can use stability-ordered cycle expansions. Stability truncation is thus easier to implement for a generic dynamical system than the curvature expansions (23.8) that rely on finite subshift approximations to a given flow.

Cycles can be detected numerically by searching a long trajectory for near recurrences. The long trajectory method for detecting cycles preferentially finds the least unstable cycles, regardless of their topological period. Another practical advantage of the method (in contrast to blind Newton method searches) is that it preferentially finds cycles in a given connected ergodic component of state space, ignoring isolated cycles or other ergodic regions elsewhere in state space.

Why should stability-ordered cycle expansions of a dynamical zeta function converge better than the crude trace formula (27.9), to be discussed in sect. 27.2? The argument has essentially already been laid out in sect. 18.6: in truncations that respect shadowing, most of the pseudo-cycles appear in shadowing combinations and nearly cancel, while only the relatively small subset affected by the increasingly long pruning rules is not shadowed. The error is typically of the order of $1/\Lambda$, which is smaller by a factor e^{hT} than the trace formula (27.9) error, where h is the entropy and T is the typical cycle period for cycles of stability Λ .

23.7.2 Smoothing



If most, but not all long cycles in a stability truncation are shadowed by shorter cycles, we say that the shadowing is partial. The breaking of exact shadowing cancellations deserves further comment. Any partial shadowing that may be present can be (partially) restored by smoothing the stability-ordered cycle expansions by replacing the $1/\Lambda$ weight for each term with the pseudo-cycle stability $\Lambda = \Lambda_{p_1} \cdots \Lambda_{p_k}$ by $f(\Lambda)/\Lambda$. Here, $f(\Lambda)$ decreases monotonically from $f(0) = 1$ to $f(\Lambda_{\max}) = 0$. The lack of smoothing means we have a step function.

A typical “shadowing error” induced by the cutoff is due to two pseudo-cycles of stability Λ separated by $\Delta\Lambda$; the contributions of these pseudo-cycles are of opposite sign. Ignoring possible weighting factors, the magnitude of the resulting term is of order $1/\Lambda - 1/(\Lambda + \Delta\Lambda) \approx \Delta\Lambda/\Lambda^2$. With smoothing, one obtains an extra term of the form $f'(\Lambda)\Delta\Lambda/\Lambda$, which we want to minimize. A reasonable guess might be to keep $f'(\Lambda)/\Lambda$ constant and as small as possible, so that

$$f(\Lambda) = 1 - \left(\frac{\Lambda}{\Lambda_{\max}} \right)^2$$

The results of a stability-ordered expansion (23.31) should always be tested for robustness by varying the cutoff $\Lambda_{\max}$. If this introduces significant variations, smoothing is probably necessary.

Résumé

A *cycle expansion* is a series representation of a dynamical zeta function, trace formula or a spectral determinant, with products in (22.11) expanded as sums over *pseudo-cycles*, which are products of the prime cycle weights t_p .

If a flow is hyperbolic and has the topology of the Smale horseshoe (a sub-shift of finite type), dynamical zeta functions are holomorphic (have only poles in the complex s plane), the spectral determinants are entire, and the spectrum of the evolution operator is discrete. The situation is considerably more reassuring than what practitioners of quantum chaos fear; there is no ‘abscissa of absolute convergence’ and no ‘entropy barrier’, the exponential proliferation of cycles is no problem, spectral determinants are entire and converge everywhere, and the topology dictates the choice of cycles to be used in cycle expansion truncations.

In this case, the basic observation is that the motion in low-dimensional dynamical systems is organized around a few *fundamental* cycles, with the cycle expansion of the Euler product

$$1/\zeta = 1 - \sum_f t_f - \sum_n \hat{c}_n,$$

regrouped into dominant *fundamental* contributions t_f and decreasing *curvature* corrections $\hat{c}_n$. The fundamental cycles t_f have no shorter approximations; they are the ‘building blocks’ of the dynamics in the sense that all longer orbits can be approximately pieced together from them. A typical curvature contribution to $\hat{c}_n$ is the *difference* of a long cycle $\{ab\}$ and its shadowing approximation by shorter cycles $\{a\}$ and $\{b\}$, as in figure 1.11:

$$t_{ab} - t_a t_b = t_{ab}(1 - t_a t_b / t_{ab})$$

Orbits that follow the same symbolic dynamics, such as $\{ab\}$ and a ‘pseudo-cycle’ $\{a\}\{b\}$, lie close to each other, have similar weights, and for increasingly long

orbits the curvature corrections fall off rapidly. Indeed, for systems that satisfy the ‘axiom A’ requirements, such as the 3-disk billiard, curvature expansions converge very well.

Once a set of the shortest cycles has been found, and the cycle periods, stabilities, and Birkhoff sums have been computed, the cycle averaging formulas such as (23.25) for the dynamical zeta function

$$\begin{aligned} \langle a \rangle &= \langle A \rangle_\zeta / \langle T \rangle_\zeta, \quad \text{where for the zeta function expansions:} \\ \langle A \rangle_\zeta &= -\frac{\partial}{\partial \beta} \frac{1}{\zeta} = \sum' A_\pi t_\pi, \quad \langle T \rangle_\zeta = \frac{\partial}{\partial s} \frac{1}{\zeta} = \sum' T_\pi t_\pi \end{aligned}$$

yield the expectation value of the observable $a(x)$, i.e., the long time average over the chaotic non-wandering set).

Commentary

Remark 23.1. Alternative Periodic Orbit Theories.

Extraordinary how potent cheap music is.

— Noel Coward



There are no ‘alternative periodic orbit theories’. There is only one ergodic theory, and periodic orbits are one aspect of it, just like there is only one quantum mechanics, and WKB is one way to gain insight into it. While the eigenfunctions of quantum evolution operators are smooth Hilbert space states, the eigenfunctions of deterministic evolution operators are highly singular, nowhere differentiable functions with support on fractal sets. The deterministic eigenstates of high-dimensional ergodic flows thus cannot be computed using the methods developed for quantum eigenstates, at least not without much further thought. The ergodic, singular ‘natural measure’ is harder (and in high-dimensional state space impossible) to construct numerically than its smooth quantum cousin, the ‘ground state’, and periodic orbits seem to be the way to do it. Were ergodic theory easy, ChaosBook.org and Gaspard monograph [15] would have been a much breezier reads.

In the vast and vastly uneven periodic orbit literature (should erroneous papers be cited?) one sometimes encounters the ‘escape-time weighting’ dreamed up by a band of men who mistook logarithm for exponential,

$$\langle a \rangle = \frac{\sum_p \tau_p a_p}{\sum_p \tau_p}, \quad \tau_p = \frac{1}{\sum_j^e \lambda_{p,j}}, \quad (23.32)$$

where the p sum goes over all known unstable periodic orbits, and sometimes also over judiciously chosen subsets of unstable equilibria. Here

$$\sum_j^e \lambda_{p,j} = \frac{1}{T_p} \ln \Lambda_p, \quad \Lambda_p = |\Lambda_{p,1} \Lambda_{p,2} \cdots \Lambda_{p,e}|, \quad (23.33)$$

T_p is the period of the prime periodic orbit p , $\Lambda_{p,j}$ is the j th Floquet multiplier (4.7), Λ_p is the product of the expanding multipliers $|\Lambda_{p,j}| > 1$, $j = 1, \dots, e$, and $\lambda_{p,j}$ ’s are the strictly positive Floquet exponents.

$\sum^e \lambda_{p,i}$ is the local escape rate from single repelling cycle p , of dimension $1/[\text{time}]$, so one may interpret its inverse τ_p as “an estimate of the mean time spent by the system in vicinity of periodic orbit” p . τ_p is the mean Lyapunov time of cycle p , that is, the mean time it takes for the density of neighboring trajectories in an arbitrarily small ball centered around a point on the trajectory to decrease by factor $1/e$.

The ‘escape-time weighting’ was introduced in a rapid communication thusly: “Less unstable orbits must be weighted more heavily, so the attractor dimension is approximated by ‘escape-time weighting’ (23.32).” That’s it: the ‘derivation’ in its entirety. Formula (23.32) is then asserted to approximate the time average $\langle a \rangle$ of observable $a(x)$ over the chaotic attractor in terms of

$$a_p = \frac{1}{T_p} \oint_0^{T_p} d\tau a(f^\tau(x_0)), \quad x_0 \in p, \quad (23.34)$$

averaged over each and every prime periodic orbit p found in any computer exploration of a dynamical system’s state space.

The enchantment with the escape-time weighting approach lies its charming simplicity. If one has a dynamical problem, and if one has a computer one has programmed to search for periodic orbits, and if the computer brings back a set of unstable periodic orbits, all one has to do is to put $\lambda_{p,j}$ and a_p into the formula (23.32), and it returns a number - let’s say $D = 9.0 \pm 0.1$, where the error one estimates somehow - which one then publishes.

The only drawback is that the ‘formula’ is wrong. (1) It comes from nowhere. (2) τ_p has dimension of [time], but a ‘weight’ should be a dimensionless number, the likelihood that an ergodic trajectory enters the neighborhood of the periodic orbit p . (3) $\sum_p$ is the sum over *all* unstable prime periodic orbits, regardless whether they belong to the ergodic component under investigation or dwell isolated in the Moon orbit. (4) The guess for the weight τ_p is clearly wrong, as any periodic orbit, no matter how long and unstable, has the comparable weight $\sum^e \lambda_{p,i}$, as long as its Lyapunovs (instability rate per unit time) are comparable; the Lyapunov time has nothing to do with the period of the particular cycle. For that there is an even more baffling fix in the literature, with the Lyapunov time in (23.32) replaced by

$$\tau_p = T_p / \sum^e \lambda_{p,i}. \quad (23.35)$$

The fix is explained as follows: “it is reasonable also to suppose that orbits with longer periods must be weighted more heavily as they are longer and should provide a greater contribution to the total sum.” That’s it: another ‘derivation’, in its entirety. No less wrong.

The *exact* weight of the unstable prime periodic orbit p (for level sum (21.6)) was given independently by Kadanoff and Tang [16] and Ozorio de Almeida and Hannay [19] in 1983. For the deterministic trace formula for flows (21.19) it is

$$\frac{T_p e^{-T_p s}}{\left| \det(\mathbf{1} - M_p) \right|} e^{\beta A_p}, \quad A_p = T_p a_p, \quad (23.36)$$

where s is the evolution operator eigenvalue, and β is an auxiliary variable. It is determined by the dimensionless Floquet *multipliers* (eigenvalues of the periodic orbit’s monodromy matrix M_p) which grow/shrink exponentially with cycle period, *not* the Osledec Lyapunov exponents or periodic orbit Floquet exponents which measure average expansion/contraction *rate per unit time*.

The *exact* cycle averaging formulas for the expectation value of the observable a , derived in chapters 19 to 22, have form

$$\langle a \rangle = \langle A \rangle_F / \langle T \rangle_F, \quad (23.37)$$

where the form of the periodic orbit sum $\langle \cdots \rangle_F$ depends on whether it is computed from the trace formula (27.15), the dynamical zeta function (23.3), or the spectral determinant (23.15). This sum is *never* of the form (23.32).

Often one cares only about the leading long-time behavior, and for long periodic orbits approximates the denominator of (23.36) by the product of the expanding multipliers Λ_p of the monodromy matrix M_p ,

$$\frac{e^{-T_p s}}{|\det(\mathbf{1} - M_p)|} \rightarrow t_p = \frac{e^{-T_p s}}{\Lambda_p}. \quad (23.38)$$

This weight seems to have been first used in 1987 by Auerbach *et al.* [4], who computed an n th order estimate $s_{(n)}$ of the leading evolution operator eigenvalue s from the sum of all periodic points j of period n

$$1 = \sum_{j \in \text{Fix}_{f^n}} t_j e^{\beta A(x_j, n)}, \quad t_j = \frac{e^{-n s_{(n)}}}{\Lambda_j}, \quad (23.39)$$

where n is discrete time, and periodic points are fixed points of the n th iterate f^n . Even as it was written, the heuristics of this paper was superseded by the exact cycle expansions, first published in 1987 Cvitanović letter [10].

Then there is in literature an ‘Alternative Periodic Orbit Theory’ so bold and breathless that one can only call it The Heresy: the conjecture is that if one looks carefully enough, there exists a *single* periodic orbit that captures all dynamical averages of a turbulent flow. This is so wrong that one is at loss what to say: The spectra of periodic orbits and the evolution operator eigenvalues are dual to each other, just as configuration and momentum Fourier spectra are dual to each other. There is no way to describe a general problem by a single Fourier mode, and there is NO such single periodic orbit. Instead, there is the well established theory that says how periodic orbits are to be used, and how many are needed to capture the hyperbolic parts of the non-wandering set to a desired accuracy. It is as elegant and systematic as Statistical Mechanics and Quantum Field Theory. Read ChaosBook.org. But who reads books nowadays?

Of course, if one picks at random a very long periodic orbit, one will get estimates as good as from an ergodic trajectory of comparable length, but then why make life hard by insisting on exact recurrence? When one starts out, The Heresy is one of the paths to enlightenment: Berry diplomatically writes “he found one orbit” in a pean to Gutzwiller [7]. Indeed, in Gutzwiller first paper (1969) on anisotropic Kepler system, the *one* periodic orbit obtained by adiabatic deformation of a Kepler ellipse yielded 10% accuracy, which was great, as in those days it was generally believed that semiclassics should be bad for the ground state. Two years later Gutzwiller invented periodic orbit theory as a tool for physicists, applied it to the full anisotropic Kepler problem, and since then there is no turning back. Similarly, Kawahara [17] computed the first Navier-Stokes periodic orbit solution embedded in turbulence, and observed that it gave rather accurate estimates of observables such as the dissipation rate.

The strangest thing about ‘Alternative Periodic Orbit Theories’ is that since introduction of zeta functions of Smale (1967), Gutzwiller (1969), Ruelle (1976) and their cycle expansions (1987) there is no need for them whatsoever. Why would one *guess* an approximate periodic orbit weight when the *exact* weight is already known? It costs exactly the

same to compute the exact spectral determinant as it costs to compute a wrong formula, both require the same periodic orbits, Floquet multipliers, periods, and cycle-averaged observables a_p . Go figure...

Remark 23.2. Pseudo-cycle expansions, cumulant expansions. Bowen's introduction of shadowing ϵ -pseudo-orbits [9] was a significant contribution to Smale's theory. The expression 'pseudo-orbits' seems to have been introduced in Parry and Pollicott's 1983 paper [20]. Following them, M. Berry [6] used the expression 'pseudo-orbits' in his 1986 paper on Riemann zeta and quantum chaology. Cycle and curvature expansions of dynamical zeta functions and spectral determinants in terms of pseudo-cycles were introduced in refs. [2, 10]. If the Möbius function (18.26) -0 if n is not square-free and is $(-1)^k$ if n is a product of k distinct primes— is used as a weight, pseudo-cycle expansion (23.1), (23.3) take form

remark A20.1

$$1/\zeta = \sum_{\pi}^{\infty} \mu(\pi) t_{\pi}, \quad (23.40)$$

known in ergodic theory as the Sarnak sum [18]

$$S_n(f)(x) = \sum_{k=1}^n \mu(k) a(f^k(x)), \quad (23.41)$$

a Möbius-weighted variant of the Birkhoff sum, see remark 20.1.

remark 20.1

Some literature [22] refers to pseudo-orbits as 'composite orbits', and to cycle expansions as 'Dirichlet series'. To statistical mechanicians, curvature expansions are very reminiscent of cumulant expansions. Indeed, (23.16) is the standard Plemelj-Smithies cumulant formula for the Fredholm determinant. A new aspect, not reminiscent of statistical mechanics, is that in cycle expansions each Q_n coefficient is expressed as a sum over exponentially many cycles.

section A20.1

Going from $N_n \approx N^n$ periodic points of period n to M_n prime cycles reduces the number of computations from N_n to $M_n \approx N^{n-1}/n$. The use of discrete symmetries (chapter 25) reduces the number of n th level terms by another factor. While reformulating theory from trace (21.24) to cycle expansion (23.8) does not eliminate exponential growth in the number of cycles, in practice only the shortest cycles are used, and the reduction in computational labor for these cycles can be significant.

Remark 23.3. Shadowing cycle-by-cycle. A glance at the low order curvatures in table 23.1 leads to the temptation to associate curvatures to individual cycles, such as $\hat{c}_{0001} = t_{0001} - t_{0+001}$. Such combinations tend to be numerically small (see, for example, ref. [3], table 1). However, splitting $\hat{c}_n$ into individual cycle curvatures is not possible in general [5]; the first example of such ambiguity in the binary cycle expansion is given by the t_{100101}, t_{100110} $0 \leftrightarrow 1$ symmetric pair of 6-cycles; the counterterm $t_{001+011}$ in table 23.1 is shared by these two cycles.

Remark 23.4. Escape rates. A lucid introduction to escape from repellers is given by Kadanoff and Tang [16]. For a review of transient chaos see refs. [8, 23]. The ζ -function formulation is given by Ruelle [21] and W. Parry and M. Pollicott [20] and discussed in ref. [10]. Altmann and Tel [1] study of escape rates in detail, with citations to literature up to 2008.

Remark 23.5. Stability ordering. The stability ordering was introduced by Dahlgvist and Russberg [11, 12] in a study of chaotic dynamics for the $(x^2y^2)^{1/a}$ potential. The presentation here runs along the lines of Dettmann and Morriss [14] for the Lorentz gas, which is hyperbolic but with highly pruned symbolic dynamics, and Dettmann and Cvitanović [13] for a family of intermittent maps. In the applications discussed in the above papers, stability ordering yields a considerable improvement over topological length ordering. In quantum chaos applications, cycle expansion cancelations are affected by the phases of pseudo-cycles (their actions), hence *period* or *action ordering* rather than stability is frequently employed.

Remark 23.6. Desymmetrized cycle expansions. The 3-disk cycle expansions (25.54) might be useful for cross-checking purposes, but, as we shall see in chapter 25, they are not recommended for actual computations, as the factorized zeta functions yield much better convergence.

References

- [1] E. G. Altmann and T. Tél, “Poincaré recurrences and transient chaos in systems with leaks”, *Phys. Rev. E* **79**, 016204 (2009).
- [2] R. Artuso, E. Aurell, and P. Cvitanović, “Recycling of strange sets: I. Cycle expansions”, *Nonlinearity* **3**, 325–359 (1990).
- [3] R. Artuso, E. Aurell, and P. Cvitanović, “Recycling of strange sets: II. Applications”, *Nonlinearity* **3**, 361–386 (1990).
- [4] D. Auerbach, P. Cvitanović, J.-P. Eckmann, G. Gunaratne, and I. Procaccia, “Exploring chaotic motion through periodic orbits”, *Phys. Rev. Lett.* **58**, 2387–2389 (1987).
- [5] E. Aurell, “Convergence of dynamical zeta functions”, *J. Stat. Phys.* **58**, 967–995 (1990).
- [6] M. V. Berry, Riemann’s Zeta function: A model for quantum chaos?, in *Quantum Chaos and Statistical Nuclear Physics*, edited by T. H. Seligman and H. Nishioka (Springer, Berlin, 1986), pp. 1–17.
- [7] M. V. Berry, “Martin Gutzwiller and his periodic orbits”, *Commun. Swiss Phys. Soc.* **37**, 34–38 (2012).
- [8] S. Bleher, C. Grebogi, and E. Ott, “Bifurcation to chaotic scattering”, *Physica D.* **46**, 87–121 (1990).
- [9] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms* (Springer, New York, 1975).
- [10] P. Cvitanović, “Invariant measurement of strange sets in terms of cycles”, *Phys. Rev. Lett.* **61**, 2729–2732 (1988).
- [11] P. Dahlgvist, “Determination of resonance spectra for bound chaotic systems”, *J. Phys. A* **27**, 763–785 (1994).
- [12] P. Dahlgvist and G. Russberg, “Periodic orbit quantization of bound chaotic systems”, *J. Phys. A* **24**, 4763–4778 (1991).

- [13] C. P. Dettmann and P. Cvitanović, “Cycle expansions for intermittent diffusion”, *Phys. Rev. E* **56**, 6687–6692 (1997).
- [14] C. P. Dettmann and G. P. Morriss, “Stability ordering; strong field Lorentz gas”, *Phys. Rev. Lett.* **78**, 4201–4204 (1997).
- [15] P. Gaspard, *Chaos, Scattering and Statistical Mechanics* (Cambridge Univ. Press, Cambridge UK, 1998).
- [16] L. Kadanoff and C. Tang, “Escape rate from strange repellers”, *Proc. Natl. Acad. Sci. USA* **81**, 1276–1279 (1984).
- [17] G. Kawahara and S. Kida, “Periodic motion embedded in plane Couette turbulence: Regeneration cycle and burst”, *J. Fluid Mech.* **449**, 291–300 (2001).
- [18] J. Liu and P. Sarnak, “The Möbius function and distal flows”, *Duke Math. J.* **164**, 1353–1399 (2015).
- [19] A. M. Ozorio de Almeida and J. H. Hannay, “Periodic orbits and a correlation function for the semiclassical density of states”, *J. Phys. A* **17**, 3429 (1984).
- [20] W. Parry and M. Pollicott, “An analogue of the prime number theorem for closed orbits of Axiom A flows”, *Ann. Math.* **118**, 573–591 (1983).
- [21] D. Ruelle, “Zeta-functions for expanding maps and Anosov flows”, *Inv. Math.* **34**, 231–242 (1976).
- [22] M. Sieber and F. Steiner, “Classical and quantum mechanics of a strongly chaotic billiard system”, *Physica D* **44**, 248–266 (1990).
- [23] T. Tél, “On the organisation of transient chaos-application to irregular scattering”, *J. Phys. A* **22**, L691–L698 (1989).

23.8 Examples

Example 23.1. Newton algorithm for determining the evolution operator eigenvalues:

 Cycle expansions of spectral determinants can be used to compute a set of leading eigenvalues of the evolution operator. A convenient way to search for these is by plotting either the absolute magnitude $\ln |\det(s - \mathcal{A})|$ or the phase of spectral determinants and dynamical zeta functions as functions of the complex variable s . The eye is guided to the zeros of spectral determinants and dynamical zeta functions by means of complex s plane contour plots, with different intervals of the absolute value of the function under investigation assigned different colors; zeros emerge as centers of elliptic neighborhoods of rapidly changing colors. Detailed scans of the whole area of the complex s plane under investigation and searches for the zeros of spectral determinants, figure 23.1, reveal complicated patterns of resonances even for something as simple as the 3-disk game of pinball. With a good starting guess (such as the location of a zero suggested by the complex s scan of figure 23.1), a zero $1/\zeta(s) = 0$ can now be determined by standard numerical methods, such as the iterative Newton algorithm (7.3), with the m th Newton estimate given by

$$s_{m+1} = s_m - \left(\zeta(s_m) \frac{\partial}{\partial s} \zeta^{-1}(s_m) \right)^{-1} = s_m - \frac{1/\zeta(s_m)}{\langle T \rangle_\zeta}. \quad (23.42)$$

The denominator $\langle T \rangle_\zeta$ is required for Newton iteration and is given by cycle expansion (23.25). We need to evaluate it anyhow, as $\langle T \rangle_\zeta$ is needed for the cycle averaging formulas.

[return to p. 456](#)

Example 23.2. Cycle expansion for the mean cycle period: For example, for the complete binary symbolic dynamics the mean cycle period $\langle T \rangle_\zeta$ is given by

[section 1.5.4](#)

$$\begin{aligned} \langle T \rangle_\zeta &= \frac{T_0}{|\Lambda_0|} + \frac{T_1}{|\Lambda_1|} + \left(\frac{T_{01}}{|\Lambda_{01}|} - \frac{T_0 + T_1}{|\Lambda_0 \Lambda_1|} \right) \\ &+ \left(\frac{T_{001}}{|\Lambda_{001}|} - \frac{T_{01} + T_0}{|\Lambda_{01} \Lambda_0|} \right) + \left(\frac{T_{011}}{|\Lambda_{011}|} - \frac{T_{01} + T_1}{|\Lambda_{01} \Lambda_1|} \right) + \dots \end{aligned} \quad (23.43)$$

Note that the cycle expansions for averages are grouped into the same shadowing combinations as the dynamical zeta function cycle expansion (23.8), with nearby pseudo-cycles nearly canceling each other.

[return to p. 460](#)

Example 23.3. Cycle expansion formula for Lyapunov exponents: In sect. 20.4 we defined the Lyapunov exponent for a 1-dimensional map, relating it to the leading eigenvalue of an evolution operator, and promised to evaluate it. Now we are finally in position to deliver on our promise.

The cycle averaging formula (23.26) yields an exact explicit expression for the Lyapunov exponent in terms of prime cycles:

$$\lambda = \frac{1}{\langle n \rangle_\zeta} \sum' (-1)^{k+1} \frac{\log |\Lambda_{p_1}| + \dots + \log |\Lambda_{p_k}|}{|\Lambda_{p_1} \dots \Lambda_{p_k}|}. \quad (23.44)$$

For a repeller, the $1/|\Lambda_p|$ weights are replaced by (27.10), the normalized measure weights $\exp(\gamma n_p)/|\Lambda_p|$, where γ is the escape rate.

[return to p. 461](#)

Exercises

- 23.1. **Cycle expansions.** Write programs that implement *binary* symbolic dynamics cycle expansions for (a) dynamical zeta functions, (b) spectral determinants. Combined with the cycles computed for a 2-branch repeller or a 3-disk system they will be useful in the problems below.
- 23.2. **Escape rate for a 1-dimensional repeller.** (continuation of exercise 20.2 - easy, but long) Consider again the quadratic map (20.38)

$$f(x) = Ax(1 - x)$$

on the unit interval. In order to be definitive, take either $A = 9/2$ or $A = 6$. Describing the itinerary of any trajectory by the binary alphabet $\{0, 1\}$ ('0' if the iterate is in the first half of the interval and '1' if it is in the second half), we have a repeller with a complete binary symbolic dynamics.

- Sketch the graph of f and determine its two fixed points $\bar{0}$ and $\bar{1}$, along with their stabilities.
- Sketch the two branches of f^{-1} . Determine all the prime cycles up to topological period 4 using your calculator and backwards iteration of f (see sect. 7.1.1).
- Determine the leading zero of the zeta function (22.11) using the weights $t_p = z^{n_p}/|\Lambda_p|$, where Λ_p is the stability of the p -cycle.
- Show that for $A = 9/2$ the escape rate of the repeller is $0.361509 \dots$ using the spectral determinant with the same cycle weight. If you have taken $A = 6$, show instead that the escape rate is in $0.83149298 \dots$. Compare the coefficients of the spectral determinant and the zeta function cycle expansions. Which expansion converges faster?

(Per Rosenqvist)

- 23.3. **Escape rate for the Ulam map.** (Medium; repeat of exercise 16.1) We will try to compute the escape rate for the Ulam map (14.22)

$$f(x) = 4x(1 - x),$$

using the method of cycle expansions. The answer should be zero, as nothing escapes.

- Compute a few of the stabilities for this map. Show that $\Lambda_0 = 4$, $\Lambda_1 = -2$, $\Lambda_{01} = -4$, $\Lambda_{001} = -8$ and $\Lambda_{011} = 8$.
- Show that

$$\Lambda_{\epsilon_1 \dots \epsilon_n} = \pm 2^n$$

and determine a rule for the sign.

- (hard) Compute the dynamical zeta function for this system

$$\zeta^{-1} = 1 - t_0 - t_1 - (t_{01} - t_0 t_1) - \dots$$

Note that the convergence as a function of the truncation cycle period is slow. Try to fix that by treating the $\Lambda_0 = 4$ cycle separately. (continued as exercise 23.11)

- 23.4. **Pinball escape rate, semi-analytical.** Estimate the 3-disk pinball escape rate for $R : a = 6$ by substituting analytical cycle stabilities and periods (see exercise 16.6 and exercise 16.7) into the appropriate binary cycle expansion. Compare your result with the numerical estimate exercise 20.3.
- 23.5. **Pinball escape rate, from numerical cycles.** Compute the escape rate for the 3-disk pinball with $R : a = 6$ by substituting the list of numerically computed cycle stabilities of exercise 16.5 into the binary cycle expansion.
- 23.6. **Pinball resonances in the complex plane.** Plot the logarithm of the absolute value of the dynamical zeta function and/or the spectral determinant cycle expansion (23.5) as contour plots in the complex s plane. Do you find zeros other than the one corresponding to the complex one? Do you see evidence for a finite radius of convergence for either cycle expansion?
- 23.7. **Counting the 3-disk pseudocycles.** (continuation of exercise 18.12) Show that the number of terms in the 3-disk pinball curvature expansion (25.53) is given by

$$\begin{aligned}
 \prod_p (1 + t_p) &= \frac{1 - 3z^4 - 2z^6}{1 - 3z^2 - 2z^3} \\
 &= 1 + 3z^2 + 2z^3 + \frac{z^4(6 + 12z + 2z^2)}{1 - 3z^2 - 2z^3} \\
 &= 1 + 3z^2 + 2z^3 + 6z^4 + 12z^5 \\
 &\quad + 20z^6 + 48z^7 + 84z^8 + 184z^9 + \dots
 \end{aligned}$$

This means that, for example, c_6 has a total of 20 terms, in agreement with the explicit 3-disk cycle expansion (25.54).

- 23.8. **4-disk unfactorized dynamical zeta function cycle expansions.** For the symmetrically arranged 4-disk pinball, the symmetry group is C_{4v} , which is of order 8. The degenerate cycles can have multiplicities 2, 4 or 8 (see table 18.3). Show that:

$$\begin{aligned}
 1/\zeta &= (1 - z^2 t_{12})^4 (1 - z^2 t_{13})^2 (1 - z^3 t_{123})^8 \\
 &\quad (1 - z^4 t_{1213})^8 (1 - z^4 t_{1214})^4 (1 - z^4 t_{1234})^2 \\
 &\quad (1 - z^4 t_{1243})^4 (1 - z^5 t_{12123})^8 (1 - z^5 t_{12124})^8 \\
 &\quad (1 - z^5 t_{12134})^8 (1 - z^5 t_{12143})^8 \\
 &\quad (1 - z^5 t_{12313})^8 (1 - z^5 t_{12413})^8 \dots
 \end{aligned} \tag{23.45}$$

Show that the cycle expansion is given by

$$\begin{aligned}
 1/\zeta &= 1 - z^2(4t_{12} + 2t_{13}) - 8z^3 t_{123} \\
 &\quad - z^4(8t_{1213} + 4t_{1214} + 2t_{1234} + 4t_{1243}) \\
 &\quad - 6t_{12}^2 - t_{13}^2 - 8t_{12}t_{13} \\
 &\quad - 8z^5(t_{12123} + t_{12124} + t_{12134} + t_{12143} + t_{12313} \\
 &\quad + t_{12413} - 4t_{12}t_{123} - 2t_{13}t_{123}) \\
 &\quad - 4z^6(2S_8 + S_4 + t_{12}^3 + 3t_{12}^2 t_{13} + t_{12}t_{13}^2 \\
 &\quad - 8t_{12}t_{1213} - 4t_{12}t_{1214} \\
 &\quad - 2t_{12}t_{1234} - 4t_{12}t_{1243} \\
 &\quad - 4t_{13}t_{1213} - 2t_{13}t_{1214} - t_{13}t_{1234} \\
 &\quad - 2t_{13}t_{1243} - 7t_{123}^2) - \dots
 \end{aligned}$$

where in the coefficient of z^6 the abbreviations S_8 and S_4 stand for the sums over the weights of the 12 orbits with multiplicity 8 and the 5 orbits with multiplicity 4, respectively; the orbits are listed in table 18.5.

- 23.9. **Escape rate for the Rössler flow.** (continuation of exercise 7.1) Try to compute the escape rate for the Rössler flow (2.30) using the method of cycle expansions. The answer should be zero, as nothing escapes. Ideally you should already have computed the cycles and have an approximate grammar, but failing that you can cheat a bit and peak into exercise 7.1.
- 23.10. **State space volume contraction, recycled.** (continuation of exercise 4.3) The plot of instantaneous state space volume contraction as a function of time in exercise 4.3 (d) illustrates one problem of time-averaging in chaotic flows - the observable might vary wildly across each recurrence to a given Poincaré section. Evaluated on a given short cycle, the average is crisp and arbitrarily accurate. Recompute $\langle \partial \cdot v \rangle$ by means of cycle expansion, study its convergence. $1/t$ convergence of mindless time-averaging is now replaced by exponential convergence in the cycle period.
- 23.11. **Ulam map is conjugate to the tent map.** (continuation of exercise 23.3, repeat of exercise A2.3 and exercise 16.2; requires real smarts, unless you look it up) Explain the magically simple form of cycle stabilities of exercise 23.3 by constructing an explicit smooth conjugacy (2.15)

$$g^t(y_0) = h \circ f^t \circ h^{-1}(y_0)$$

that conjugates the Ulam map (14.22) into the tent map (14.21).

- 23.12. **Continuous vs. discrete mean return time.** Show that the expectation value $\langle a \rangle$ time-averaged over continuous time flow is related to the corresponding average $\langle a \rangle_{\text{dscr}}$ measured in discrete time (e.g., Poincaré section returns) by (23.28):

$$\langle a \rangle_{\text{dscr}} = \langle a \rangle \langle T \rangle_{\zeta} / \langle n \rangle_{\zeta}. \quad (23.46)$$

(Hint: consider the form of their cycle expansions.) The mean discrete period $\langle n \rangle_{\zeta}$ averaged over cycles, and the mean continuous time period $\langle T \rangle_{\zeta}$ need to be evaluated only once, thereafter one can compute either $\langle a \rangle$ or $\langle a \rangle_{\text{dscr}}$, whichever is more convenient.

Chapter 24

Deterministic diffusion

This is a bizzare and discordant situation.

—M.V. Berry

(R. Artuso and P. Cvitanović)

THE ADVANCES in the theory of dynamical systems have brought a new life to Boltzmann's mechanical formulation of statistical mechanics. Sinai, Ruelle and Bowen (SRB) have generalized Boltzmann's notion of ergodicity for a constant energy surface for a Hamiltonian system in equilibrium to dissipative systems in nonequilibrium stationary states. In this more general setting the attractor plays the role of a constant energy surface, and the SRB measure of sect. 19.1 is a generalization of the Liouville measure. Such measures are purely microscopic and indifferent to whether the system is at equilibrium, close to equilibrium or far from it. "Far from equilibrium" in this context refers to systems with large deviations from Maxwell's equilibrium velocity distribution. Furthermore, the theory of dynamical systems has yielded new sets of microscopic dynamics formulas for macroscopic observables such as diffusion constants and the pressure, to which we turn now.



We shall apply cycle expansions to the analysis of *transport* properties of chaotic systems.

The resulting formulas are exact; no probabilistic assumptions are made, and the all correlations are taken into account by the inclusion of cycles of all periods. The infinite extent systems for which the periodic orbit theory yields formulas for diffusion and other transport coefficients are spatially periodic, the global state space being tiled with copies of a elementary cell. The motivation are physical problems such as beam defocusing in particle accelerators or chaotic behavior of passive tracers in 2-dimensional rotating flows, problems which can be described as deterministic diffusion in periodic arrays.



In sect. 24.1 we derive the formulas for diffusion coefficients in a simple physical setting, the 2-dimensional periodic Lorentz gas. This system, however, is not

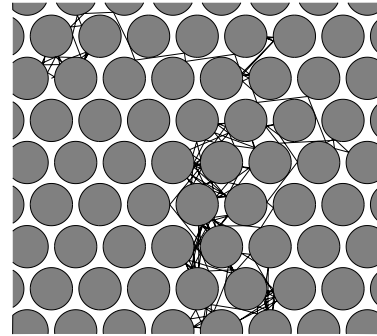


Figure 24.1: Deterministic diffusion in a finite horizon periodic Lorentz gas. (T. Zhang)

the best one to illustrate the theory, due to its complicated symbolic dynamics. Therefore we apply the theory first to diffusion induced by a 1-dimensional maps in sect. 24.2.

24.1 Diffusion in periodic arrays

Chaos happens - let's make a better use of it.
— Edward Tenner

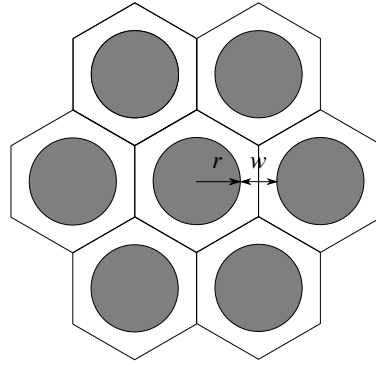
The 2-dimensional *Lorentz gas* is an infinite scatterer array in which diffusion of a light molecule in a gas of heavy scatterers is modeled by the motion of a point particle in a plane bouncing off an array of reflecting disks. The Lorentz gas is called “gas” as one can equivalently think of it as consisting of any number of pointlike fast “light molecules” interacting only with the stationary “heavy molecules” and not among themselves. As the scatterer array is built up from only defocusing concave surfaces, it is a pure hyperbolic system, and one of the simplest non-trivial dynamical systems that exhibits deterministic diffusion, figure 24.1. We shall now show that the *periodic* Lorentz gas is amenable to a purely deterministic treatment. In this class of open dynamical systems quantities characterizing global dynamics, such as the Lyapunov exponent, pressure and diffusion constant, can be computed from the dynamics restricted to the elementary cell. The method applies to any hyperbolic dynamical system that is a periodic tiling $\hat{\mathcal{M}} = \bigcup_{\hat{n} \in T} \mathcal{M}_{\hat{n}}$ of the dynamical state space $\hat{\mathcal{M}}$ by translates $\mathcal{M}_{\hat{n}}$ of an *elementary cell* $\mathcal{M}$, with T the abelian group of lattice translations (see figure 24.2). If the scattering array has further discrete rotational and reflection symmetries (G/T is a point group), each elementary cell may be built from a *fundamental domain* $\tilde{\mathcal{M}}$ by the action of a discrete (not necessarily abelian) group G . The symbol $\hat{\mathcal{M}}$ refers here to the full state space, i.e., both the spatial coordinates and the momenta. The spatial component of $\hat{\mathcal{M}}$ is the complement of the disks in the *whole* space.

remark 24.1

We shall now relate the dynamics in $\mathcal{M}$ to diffusive properties of the Lorentz gas in $\hat{\mathcal{M}}$.

These concepts are best illustrated by a specific example, a Lorentz gas based on the hexagonal lattice Sinai billiard of figure 24.3.

Figure 24.2: An elementary cell and its six nearest neighbor translations. The ratio (24.1) of the distance between a nearest pair of disks and the disk radius determines the dynamical properties in the system: the horizon is finite for $w/r \leq 0.3094 \dots$, and infinite beyond that. (T. Zhang)



We distinguish two types of diffusive behavior; the *infinite horizon* case, which allows for infinite length flights, and the *finite horizon* case, where any free particle trajectory must hit a disk in finite time. Consider figure 24.2, where w is the distance between a nearest pair of disks and r is the disk radius (here set to $r = 1$). The ratio w/r is the only parameter of the problem, the parameter that determines the dynamical properties in the system: the horizon is finite for

$$\frac{w}{r} < \frac{4}{\sqrt{3}} - 2 = 0.3094 \dots, \quad (24.1)$$

and infinite beyond that. In this chapter we shall restrict our consideration to the finite horizon case, with disks sufficiently large so that no infinite length free flight is possible. In this case the diffusion is normal, with $\hat{x}(t)^2$ growing like t . We shall discuss the anomalous diffusion case in sect. 24.3.

As we will work with three kinds of state spaces, good manners require that we repeat what tildes, nothings and hats atop symbols signify:

- $\sim$ fundamental domain, triangle in figure 24.3
 - elementary cell, hexagon in figure 24.3
 - $\wedge$ full state space, lattice in figure 24.3
- (24.2)

It is convenient to define an evolution operator for each of the 3 cases of figure 24.3. $\hat{x}(t) = \hat{f}^t(\hat{x})$ denotes the point in the global space $\hat{\mathcal{M}}$ reached by the flow in time t . $x(t) = f^t(x_0)$ denotes the corresponding flow in the elementary cell; the two are related by

$$\hat{n}_t(x_0) = \hat{f}^t(x_0) - f^t(x_0) \in T, \quad (24.3)$$

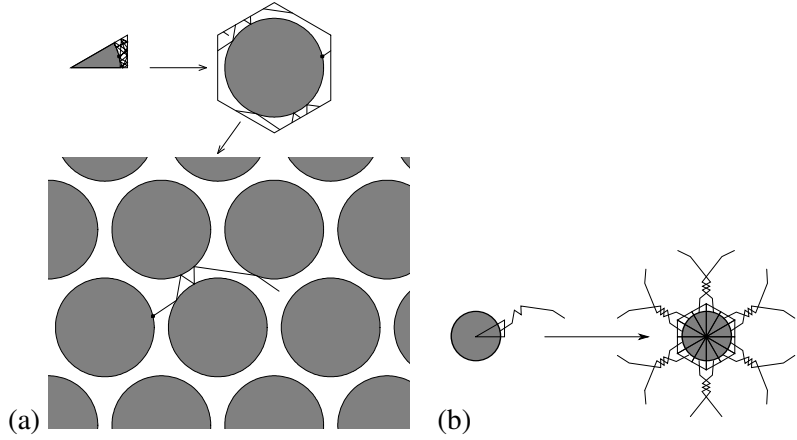
the translation of the endpoint of the global path into the elementary cell $\mathcal{M}$. The quantity $\tilde{x}(t) = \tilde{f}^t(\tilde{x})$ denotes the flow in the fundamental domain $\tilde{\mathcal{M}}$; $\tilde{f}^t(\tilde{x})$ is related to $f^t(\tilde{x})$ by a discrete symmetry $g \in G$ which maps $\tilde{x}(t) \in \tilde{\mathcal{M}}$ to $x(t) \in \mathcal{M}$.

chapter 25

Fix a vector $\beta \in \mathbb{R}^d$, where d is the dimension of the state space. We will compute the diffusive properties of the Lorentz gas from the leading eigenvalue of the evolution operator (20.10)

$$s(\beta) = \lim_{t \rightarrow \infty} \frac{1}{t} \log \langle e^{\beta \cdot (\hat{x}(t) - x)} \rangle_{\mathcal{M}}, \quad (24.4)$$

Figure 24.3: A periodic lattice of reflecting disks $\hat{\mathcal{M}}$ is tiled by copies of the fundamental domain $\tilde{\mathcal{M}}$. (a) Motion in the fundamental domain $\tilde{\mathcal{M}}$ (top left), the elementary cell $\mathcal{M}$ (top right), and the full state space $\hat{\mathcal{M}}$ (bottom). (b) The above trajectory unwrapped in the full space and its 11 copies obtained by applying the twelve D_6 point group actions to it. (from ref. [15]).



where the average is over all initial points in the elementary cell, $x \in \mathcal{M}$.

If all odd derivatives vanish by symmetry, there is no drift and the second derivatives

$$2dD_{ij} = \left. \frac{\partial}{\partial \beta_i} \frac{\partial}{\partial \beta_j} s(\beta) \right|_{\beta=0} = \lim_{t \rightarrow \infty} \frac{1}{t} \langle (\hat{x}(t) - x)_i (\hat{x}(t) - x)_j \rangle_{\mathcal{M}},$$

yield a diffusion matrix. This symmetric matrix can, in general, be anisotropic (i.e., have d distinct eigenvalues and eigenvectors). The spatial diffusion constant is then given by the Einstein relation (20.30)

$$D = \frac{1}{2d} \sum_i \left. \frac{\partial^2}{\partial \beta_i^2} s(\beta) \right|_{\beta=0} = \lim_{t \rightarrow \infty} \frac{1}{2dt} \langle (\hat{q}(t) - q)^2 \rangle_{\mathcal{M}},$$

where the i sum is restricted to the spatial components q_i of the state space vectors $x = (q, p)$, i.e., if the dynamics is Hamiltonian, the sum is over the d degrees of freedom.

We now turn to the connection between (24.4) and periodic orbits in the elementary cell. As the full $\hat{\mathcal{M}} \rightarrow \tilde{\mathcal{M}}$ reduction is complicated by the non-abelian nature of G , we discuss only the abelian $\hat{\mathcal{M}} \rightarrow \mathcal{M}$ reduction.

remark 24.5

24.1.1 Reduction from $\hat{\mathcal{M}}$ to $\mathcal{M}$

The key idea follows from inspection of the relation

$$\langle e^{\beta \cdot (\hat{x}(t) - x)} \rangle_{\mathcal{M}} = \frac{1}{|\mathcal{M}|} \int_{\substack{x \in \mathcal{M} \\ \hat{y} \in \hat{\mathcal{M}}}} dx d\hat{y} e^{\beta \cdot (\hat{y} - x)} \delta(\hat{y} - \hat{f}^t(x)).$$

$|\mathcal{M}| = \int_{\mathcal{M}} dx$ is the volume of the elementary cell $\mathcal{M}$. Due to translational symmetry, it suffices to start with a density of trajectories defined over a single elementary cell $\mathcal{M}$. As in sect. 20.2, we have used the identity $1 = \int_{\mathcal{M}} dy \delta(y - \hat{x}(t))$ to motivate the introduction of the evolution operator $\mathcal{L}^t(\hat{y}, x)$. There is a unique lattice

translation $\hat{n}$ such that $\hat{y} = y - \hat{n}$, with the endpoint $y \in \mathcal{M}$ translated back to the elementary cell, and $f^t(x)$ given by (24.3). The difference is a translation by a constant lattice vector $\hat{n}$, and the Jacobian for changing integration from $d\hat{y}$ to dy equals unity. Therefore, and this is the main point, translation invariance can be used to reduce this average to the elementary cell:

$$\langle e^{\beta \cdot (\hat{x}(t) - x)} \rangle_{\mathcal{M}} = \frac{1}{|\mathcal{M}|} \int_{x,y \in \mathcal{M}} dx dy e^{\beta \cdot (\hat{f}^t(x) - x)} \delta(y - f^t(x)). \quad (24.5)$$

As this is a translation, the Jacobian is $|\partial \hat{y} / \partial y| = 1$. In this way the global $\hat{f}^t(x)$ flow, infinite volume state space averages can be computed by following the flow $f^t(x_0)$ restricted to the compact, finite volume elementary cell $\mathcal{M}$. The equation (24.5) suggests that we study the evolution operator

$$\mathcal{L}^t(y, x) = e^{\beta \cdot (\hat{x}(t) - x)} \delta(y - f^t(x)), \quad (24.6)$$

where $\hat{x}(t) = \hat{f}^t(x) \in \hat{\mathcal{M}}$ is the displacement in the full space, but $x, f^t(x), y \in \mathcal{M}$. It is straightforward to check that this operator satisfies the semigroup property (20.16),

$$\int_{\mathcal{M}} dz \mathcal{L}^{t_2}(y, z) \mathcal{L}^{t_1}(z, x) = \mathcal{L}^{t_2+t_1}(y, x).$$

For $\beta = 0$, the operator (24.6) is the Perron-Frobenius operator (19.10), with the leading eigenvalue $e^{s_0} = 1$ because there is no escape from this system (see the flow conservation sum rule (23.17)).

The rest is old hat. The spectrum of $\mathcal{L}$ is evaluated by taking the trace

$$\text{tr } \mathcal{L}^t = \int_{\mathcal{M}} dx e^{\beta \cdot \hat{n}_t(x)} \delta(x - x(t)).$$

Here $\hat{n}_t(x)$ is the discrete lattice translation defined in (24.3). Two kinds of orbits periodic in the elementary cell contribute. A periodic orbit is called *standing* if it is also periodic orbit of the infinite state space dynamics, $\hat{f}^{T_p}(x) = x$, and it is called *running* if it corresponds to a lattice translation in the dynamics on the infinite state space, $\hat{f}^{T_p}(x) = x + \hat{n}_p$. We recognize the shortest repeating segment of a running orbit as our old ‘relative periodic orbit’ friend from chapter 11. In the theory of area-preserving maps such as the standard map of example 8.7 these orbits are called *accelerator modes*, as the diffusion takes place along the momentum rather than the position coordinate. The traveled distance $\hat{n}_p = \hat{n}_{T_p}(x_0)$ is independent of the starting point x_0 , as can be easily seen by continuing the path periodically in $\hat{\mathcal{M}}$.

The final result is the spectral determinant (22.5)

$$\det(s(\beta) - \mathcal{A}) = \prod_p \exp \left(- \sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{(\beta \cdot \hat{n}_p - s T_p)r}}{|\det(\mathbf{1} - M_p^r)|} \right), \quad (24.7)$$

or the corresponding dynamical zeta function (22.11)

$$1/\zeta(\beta, s) = \prod_p \left(1 - \frac{e^{(\beta \cdot \hat{n}_p - s T_p)}}{|\Lambda_p|} \right). \quad (24.8)$$

The dynamical zeta function cycle averaging formula (23.24) for the diffusion constant (20.30), zero mean drift $\langle \hat{x}_i \rangle = 0$, is given by



$$D = \frac{1}{2d} \frac{\langle \hat{x}^2 \rangle_\zeta}{\langle T \rangle_\zeta} = \frac{1}{2d} \frac{1}{\langle T \rangle_\zeta} \sum' \frac{(-1)^{k+1} (\hat{n}_{p_1} + \dots + \hat{n}_{p_k})^2}{|\Lambda_{p_1} \dots \Lambda_{p_k}|}. \quad (24.9)$$

where the sum is over all distinct non-repeating combination of prime cycles. The derivation is standard, still the formula is strange. Diffusion is unbounded motion across an infinite lattice; nevertheless, the reduction to the elementary cell enables us to compute relevant quantities in the usual way, in terms of periodic orbits.



A sleepy reader might protest that $x(T_p) - x(0)$ is manifestly equal to zero for a periodic orbit. That is correct; $\hat{n}_p$ in the above formula refers to a displacement $\hat{x}(T_p)$ on the *infinite* periodic lattice, while p refers to closed orbit of the dynamics $f^t(x)$ reduced to the elementary cell, with x_p a periodic point in the closed prime cycle p .

Even so, this is not an obvious formula. Globally periodic orbits have $\hat{x}_p^2 = 0$, and contribute only to the time normalization $\langle T \rangle_\zeta$. The mean square displacement $\langle \hat{x}^2 \rangle_\zeta$ gets contributions only from the periodic runaway trajectories; they are closed in the elementary cell, but on the periodic lattice each one grows like $\hat{x}(t)^2 = (\hat{n}_p/T_p)^2 t^2 = v_p^2 t^2$. So the orbits that contribute to the trace formulas and spectral determinants exhibit either ballistic transport or no transport at all: diffusion arises as a balance between the two kinds of motion, weighted by the $1/|\Lambda_p|$ measure. If the system is not hyperbolic such weights may be abnormally large, with $1/|\Lambda_p| \approx 1/T_p^\alpha$ rather than $1/|\Lambda_p| \approx e^{-T_p \lambda}$, where λ is the Lyapunov exponent, and they may lead to anomalous diffusion - accelerated or slowed down depending on whether the probabilities of the running or the standing orbits are enhanced.



section 24.3

We illustrate the main idea, tracking of a globally diffusing orbit by the associated confined orbit restricted to the elementary cell, with a class of simple 1-dimensional dynamical systems where all transport coefficients can be evaluated analytically. For another example of deterministic diffusion in a Hamiltonian system, consult appendix A24.

appendix A24

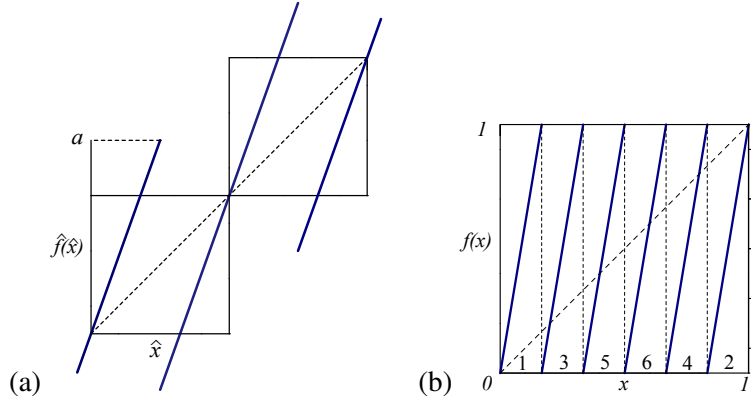
24.2 Diffusion induced by chains of 1-dimensional maps

In a typical deterministic diffusive process, trajectories originating from a given scatterer reach a finite set of neighboring scatterers in one bounce, and then the process is repeated. As was shown in chapter 14, the essential part of this process is the stretching along the unstable directions of the flow, and in the crudest approximation the dynamics can be modeled by 1-dimensional expanding maps. This observation motivates introduction of a class of particularly simple 1-dimensional systems.



example 24.1
p. 492

Figure 24.4: (a) $\hat{f}(\hat{x})$, the full space sawtooth map (24.21), $\Lambda > 2$. (b) $f(x)$, the sawtooth map restricted to the unit circle (24.24), $\Lambda = 6$.



As noted in sect. 24.1.1, the elementary cell cycles correspond to either standing or running orbits for the map on the full line: we shall refer to $\hat{n}_p \in \mathbb{Z}$ as the *jumping number* of the p cycle, and take as the cycle weight

$$t_p = z^{n_p} e^{\beta \hat{n}_p} / |\Lambda_p|. \quad (24.10)$$

The diffusion constant formula (24.9) for 1-dimensional maps is

$$D = \frac{1}{2} \frac{\langle \hat{n}^2 \rangle_\zeta}{\langle n \rangle_\zeta}, \quad (24.11)$$

where the “mean cycle time” is given by (23.25)

$$\langle n \rangle_\zeta = z \frac{\partial}{\partial z} \frac{1}{\zeta(0, z)} \Big|_{z=1} = - \sum' (-1)^k \frac{n_{p_1} + \cdots + n_{p_k}}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}, \quad (24.12)$$

and the “mean cycle displacement squared” by (23.27)

$$\langle \hat{n}^2 \rangle_\zeta = \frac{\partial^2}{\partial \beta^2} \frac{1}{\zeta(\beta, 1)} \Big|_{\beta=0} = - \sum' (-1)^k \frac{(\hat{n}_{p_1} + \cdots + \hat{n}_{p_k})^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}, \quad (24.13)$$

the primed sum indicating all distinct non-repeating combinations of prime cycles. The evaluation of these formulas for the simple system of example 24.1 will require nothing more than pencil and paper.

remark 24.4



example 24.2
p. 492

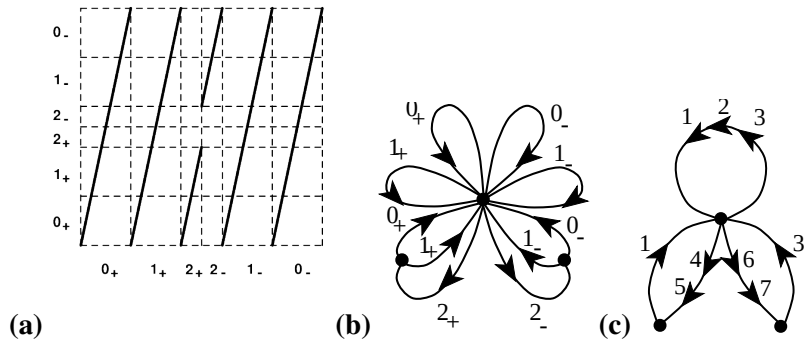
24.2.1 Higher order transport coefficients

The same approach yields higher order transport coefficients

$$\mathcal{B}_k = \frac{1}{k!} \frac{d^k}{d\beta^k} s(\beta) \Big|_{\beta=0}, \quad \mathcal{B}_2 = D, \quad (24.14)$$

known for $k > 2$ as the Burnett coefficients. The behavior of the higher order coefficients yields information on the relaxation to the asymptotic distribution function

Figure 24.5: (a) A partition of the unit interval into six intervals, example 24.4, labeled by the jumping number $\hat{n}(x)$ $I = \{0_+, 1_+, 2_+, 2_-, 1_-, 0_-\}$. The partition is Markov, as the critical point is mapped onto the right border of M_{1_+} . (b) The transition graph for this partition. (c) The transition graph in the compact Vadim Moroz notation of (24.32).



generated by the diffusive process. Here $\hat{x}_t$ is the relevant dynamical variable and $\mathcal{B}_k$'s are related to moments $\langle \hat{x}_t^k \rangle$ of arbitrary order.

Were the diffusive process purely Gaussian

$$e^{ts(\beta)} = \frac{1}{\sqrt{4\pi Dt}} \int_{-\infty}^{+\infty} d\hat{x} e^{\beta \hat{x}} e^{-\hat{x}^2/(4Dt)} = e^{\beta^2 Dt} \quad (24.15)$$

the only $\mathcal{B}_k$ coefficient different from zero would be $\mathcal{B}_2 = D$. Hence, nonvanishing higher order coefficients signal deviations of deterministic diffusion from a Gaussian stochastic process.

 example 24.3
p. 493

We see that deterministic diffusion is not a Gaussian stochastic process. Higher order even coefficients may be calculated along the same lines.

24.2.2 Finite Markov partitions

For piecewise-linear maps exact results may be obtained whenever the critical points are mapped in finite numbers of iterations onto partition boundary points, or onto unstable periodic orbits. We will work out here an example for which this occurs in two iterations, leaving other cases as exercises. The key idea is to construct a *Markov partition* (14.2), with intervals mapped *onto* unions of intervals.

 example 24.4
p. 493

It is by now clear how to build an infinite hierarchy of finite Markov partitions: tune the slope in such a way that the critical value $f(1/2)$ is mapped into the fixed point at the origin, $f^n(1/2) = 0$, in a finite number of iterations n . By taking higher and higher values of n one constructs a dense set of Markov parameter values, organized into a hierarchy that resembles the way in which rationals are densely embedded in the unit interval. For example, each of the 6 primary intervals can be subdivided into 6 intervals obtained by the 2-nd iterate of the map,

and for the critical point mapping into any of those in 2 steps the grammar (and the corresponding cycle expansion) is finite. So, if we can prove continuity of $D = D(\Lambda)$, we can apply the periodic orbit theory to the sawtooth map (24.21) for a random “generic” value of the parameter Λ , for example $\Lambda = 4.5$. The idea is to bracket this value of Λ by a sequence of nearby Markov values, compute the exact diffusion constant for each such Markov partition, and study their convergence toward the value of D for $\Lambda = 4.5$. Some details of how this is accomplished are given in appendix A14.3 for a related problem, the pruned Bernoulli shift. Judging how difficult such problem is already for a tent map (see sect. 18.5), this is not likely to take only a week of work.

Expressions like (24.28) may lead to an expectation that the diffusion coefficient (and thus transport properties) are smooth functions of parameters controlling the chaoticity of the system. For example, one might expect that the diffusion coefficient increases smoothly and monotonically as the slope Λ of the map (24.21) is increased, or, perhaps more physically, that the diffusion coefficient is a smooth function of the Lyapunov exponent λ . This turns out not to be true: D as a function of Λ is a fractal, nowhere differentiable curve illustrated in figure 24.6. The dependence of D on the map parameter Λ is rather unexpected - even though for larger Λ more points are mapped outside the unit cell in one iteration, the diffusion constant does not necessarily grow.

remark 24.3

This is a consequence of the lack of structural stability, even of purely hyperbolic systems such as the Lozi map and the 1-dimensional diffusion map (24.21). The trouble arises due to non-smooth dependence of the topological entropy on system parameters - any parameter change, no matter how small, leads to creation and destruction of infinitely many periodic orbits. As far as diffusion is concerned this means that even though local expansion rate is a smooth function of Λ , the number of ways in which the trajectory can re-enter the initial cell is an irregular function of Λ .

The lesson is that lack of structural stability implies lack of spectral stability, and no global observable is expected to depend smoothly on the system parameters. If you want to master the material, working through one of the deterministic diffusion projects on ChaosBook.org/pages is strongly recommended.

24.3 Marginal stability and anomalous diffusion

What effect does the intermittency of chapter 29 have on transport properties? A marginal fixed point affects the balance between the running and standing orbits, thus generating a mechanism that may result in anomalous diffusion.

remark 24.6



example 24.5
p. 495

D vanishes by the implicit function theorem, $z''(\beta)|_{\beta=1} = 0$ when $\alpha \leq 1$. The physical interpretation is that a typical orbit will stick for long times near the $\bar{0}$

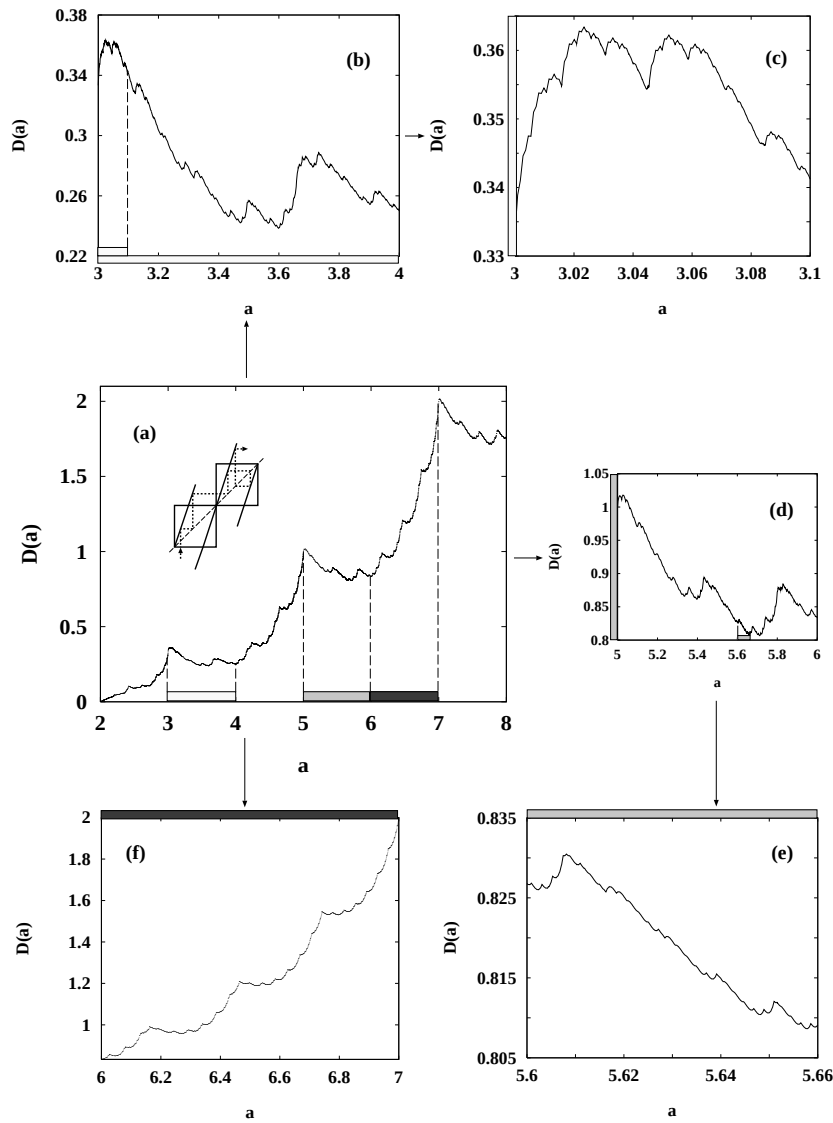


Figure 24.6: The dependence of D on the map parameter a is continuous, but not monotone. Here a stands for the slope Λ in (24.21) (from R. Klages thesis [29]).

marginal fixed point, and the ‘trapping time’ will be larger for higher values of the intermittency parameter s (recall $\alpha = 1/s$). In that case one needs to look more closely at the behavior of traces of high powers of the transfer operator.

The evaluation of transport coefficient requires one more derivative with respect to expectation values of state space observables (see sect. 24.1): if we use the diffusion dynamical zeta function (24.8), we may write the diffusion coefficient as an inverse Laplace transform, in such a way that the distinction between maps and flows has vanished. In the case of 1-dimensional diffusion we thus have

$$D = \lim_{t \rightarrow \infty} \frac{d^2}{d\beta^2} \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} ds e^{st} \frac{\zeta'(\beta, s)}{\zeta(\beta, s)} \Big|_{\beta=0} \quad (24.16)$$

where the ζ' refers to the derivative with respect to s .

The evaluation of inverse Laplace transforms for high values of the argument is most conveniently performed using Tauberian theorems. We shall take

$$\omega(\lambda) = \int_0^\infty dx e^{-\lambda x} u(x),$$

with $u(x)$ monotone as $x \rightarrow \infty$; then, as $\lambda \mapsto 0$ and $x \mapsto \infty$ respectively (and $\rho \in (0, \infty)$),

$$\omega(\lambda) \sim \frac{1}{\lambda^\rho} L\left(\frac{1}{\lambda}\right)$$

if and only if

$$u(x) \sim \frac{1}{\Gamma(\rho)} x^{\rho-1} L(x),$$

where L denotes any slowly varying function with $\lim_{t \rightarrow \infty} L(ty)/L(t) = 1$. Now

$$\frac{1/\zeta_0'(e^{-s}, \beta)}{1/\zeta_0(e^{-s}, \beta)} = \frac{\left(\frac{4}{\Lambda} + \frac{\Lambda-4}{\Lambda\zeta(1+\alpha)} (J(e^{-s}, \alpha+1) + J(e^{-s}, \alpha))\right) \cosh \beta}{1 - \frac{4}{\Lambda} e^{-s} \cosh \beta - \frac{\Lambda-4}{\Lambda\zeta(1+\alpha)} e^{-s} (e^{-s}, \alpha+1) \cosh \beta J}.$$

Taking the second derivative with respect to β we obtain

$$\begin{aligned} & \frac{d^2}{d\beta^2} \left(1/\zeta_0'(e^{-s}, \beta) / \zeta^{-1}(e^{-s}, \beta) \right)_{\beta=0} \\ &= \frac{\frac{4}{\Lambda} + \frac{\Lambda-4}{\Lambda\zeta(1+\alpha)} (J(e^{-s}, \alpha+1) + J(e^{-s}, \alpha))}{\left(1 - \frac{4}{\Lambda} e^{-s} - \frac{\Lambda-4}{\Lambda\zeta(1+\alpha)} e^{-s} J(e^{-s}, \alpha+1)\right)^2} = g_\alpha(s). \end{aligned} \quad (24.17)$$

The asymptotic behavior of the inverse Laplace transform (24.16) may then be evaluated via Tauberian theorems, once we use our estimate for the behavior of Jonquière functions near $z = 1$. The deviations from normal behavior correspond

to an explicit dependence of D on time. Omitting prefactors (which can be calculated by the same procedure) we have

$$g_\alpha(s) \sim \begin{cases} s^{-2} & \text{for } \alpha > 1 \\ s^{-(\alpha+1)} & \text{for } \alpha \in (0, 1) \\ 1/(s^2 \ln s) & \text{for } \alpha = 1. \end{cases}$$

The anomalous diffusion exponents follow:

exercise 24.6

$$\langle (x - x_0)^2 \rangle_t \sim \begin{cases} t^\alpha & \text{for } \alpha \in (0, 1) \\ t / \ln t & \text{for } \alpha = 1 \\ t & \text{for } \alpha > 1. \end{cases} \quad (24.18)$$

Résumé

Perfection itself is imperfection.

— Vladimir Horowitz

With initial data accuracy $\delta x = |\delta \mathbf{x}(0)|$ and system size L , a trajectory is predictable only to the finite Lyapunov time $T_{\text{Lyap}} \approx \lambda^{-1} \ln |L/\delta x|$. Beyond that, chaos rules. We have discussed the implications in sect. 1.8: chaos is good news for prediction of long term observables such as transport in statistical mechanics.

The classical Boltzmann equation for evolution of 1-particle density is based on *stosszahlansatz*, neglect of particle correlations prior to, or after a 2-particle collision. It is a very good approximate description of dilute gas dynamics, but a difficult starting point for inclusion of systematic corrections. In the theory developed here, no correlations are neglected - they are all included in the cycle averaging formula such as the cycle expansion for the diffusion constant

$$D = \frac{1}{2d} \frac{1}{\langle T \rangle_\zeta} \sum' (-1)^{k+1} \frac{(\hat{n}_{p_1} + \cdots + \hat{n}_{p_k})^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}.$$

Such formulas are *exact*; the issue in their applications is what are the most effective schemes of estimating the infinite cycle sums required for their evaluation. Here there are no phenomenological macroscopic parameters; quantities such as transport coefficients are calculable to any desired accuracy from the microscopic dynamics.

For systems of a few degrees of freedom these results are on rigorous footing, but there are indications that they capture the essential dynamics of systems of many degrees of freedom as well.

Though superficially indistinguishable from the probabilistic random walk diffusion, deterministic diffusion is quite recognizable, at least in low dimensional settings, through fractal dependence of the diffusion constant on the system parameters (see sect. 1.8), and through non-Gaussian relaxation to equilibrium (non-vanishing Burnett coefficients).

section 1.8

That Smale's "structural stability" conjecture turned out to be wrong is not a bane of chaotic dynamics - it is actually a virtue, perhaps its most dramatic experimentally measurable prediction. As long as microscopic periodicity of the physical system, such as a face of a crystal, is exact, the prediction is counterintuitive for a physicist - transport coefficients are *not* smooth functions of system parameters, rather they are non-monotonic, *nowhere differentiable* functions.

Commentary

Remark 24.1. Lorentz gas. The "Lorentz gas" is one of the simplest hyperbolic Hamiltonian dynamical systems that exhibits chaos and deterministic diffusion. The original Lorentz gas [36] assumed a random distribution of heavy scatterers; a description of such gas requires statistical assumptions about the distribution of scatterers. A periodic Lorentz gas (configuration of scatterers invariant under a discrete group of translations of the plane), however, is amenable to pure deterministic description. Ergodic properties of periodic Lorentz gases were first studied by Sinai [39], and its diffusive properties have been extensively studied ever since [8–11, 19, 22, 37]. One distinguishes the *infinite horizon* diffusive behavior, which allows for infinite length flights, from the *finite horizon* case [10], where the particle always hits the next disk in finite time, and the diffusion is normal [7, 10], with $\hat{x}(t)^2$ growing like t . Most of the periodic Lorentz gas literature, such as Bunimovich and Sinai [10], is focused on the symmetries under discrete translations of periodic tilings of the plane, usually defined by a parallelepipedal "primitive unit cell" (also called "fundamental domain" in literature; here that term will be reserved for the smallest tile that tiles the hexagon). However, for a triangular periodic Lorentz gas the full symmetry group is the space group $p6mm$ (see Chapter 11 of ref. [13] for a discussion of the geometry of space groups), and the natural tiling is in terms of the hexagon centered on the scattering disk ("Wigner-Seitz cell", "Voronoi cell"). See Dettmann [16] review.



remark 11.3

Remark 24.2. Who's dunnit? Cycle expansions for the diffusion constant of a particle moving in a periodic array have been introduced by Artuso [1] (exact dynamical zeta function for 1-dimensional chains of maps (24.9)), by Vance [40] (who applied the Artuso [1] formula to the Lorentz gas) and by Cvitanović, Eckmann and Gaspard [14] (the dynamical zeta function cycle expansion (24.9) applied to the Lorentz gas). Attempts to evaluate the Lorentz gas dynamical zeta function cycle expansion were carried out by Schreiber [15] and Zhang [41].

Remark 24.3. Lack of structural stability for D . Expressions like (24.28) may lead to an expectation that the diffusion coefficient (and thus transport properties) are smooth functions of the chaoticity of the system (parameterized, for example, by the Lyapunov exponent $\lambda = \ln \Lambda$). This turns out not to be true: D as a function of Λ is a fractal, nowhere differentiable curve shown in figure 24.6. The dependence of D on the map parameter Λ is rather unexpected - even though for larger Λ more points are mapped outside the unit cell in one iteration, the diffusion constant does not necessarily grow. We refer the reader to refs. [23, 38] for early work on the deterministic diffusion induced by 1-dimensional maps. The sawtooth map (24.21) was introduced by Grossmann and Fujisaka [26] who derived the integer slope formulas (24.28) for the diffusion constant. The sawtooth map is also discussed in ref. [21]. The fractal dependence of diffusion constant on the map

parameter is discussed in refs. [29, 32, 33]. Keller, Howard and Klages [28] show that for piecewise C^2 expanding interval maps the diffusion coefficient D is Lipschitz continuous under parameter variations, up to quadratic logarithmic corrections. R. Klages lecture notes [31] offer a quick, first-year Ph.D. introduction to the concept of deterministic diffusion. For the state of the art of fractal transport coefficients, as of 2007, consult the first part of Klages's monograph [30]. Sect. 1.8 discusses briefly the experimental implications; we find this "The deepest thing about chaos". It would be sweet if someone actually checked this kind of predictions in an experiment. No fractal-like behavior of the conductivity for the Lorentz gas has been detected so far [35]. Statistical mechanics (see, for example, Gallavotti and Cohen [20]) tend to believe that such complicated behavior is not to be expected in systems with very many degrees of freedom, as the addition to a large integer dimension of a number smaller than 1 should be as unnoticeable as a microscopic perturbation of a macroscopic quantity.



Remark 24.4. Symmetry factorization in one dimension. In the $\beta = 0$ limit the dynamics (24.23) is symmetric under $x \rightarrow -x$, and the zeta functions factorize into products of zeta functions for the symmetric and antisymmetric subspaces, as described in example 25.9:

$$\begin{aligned} \frac{1}{\zeta(0, z)} &= \frac{1}{\zeta_s(0, z)} \frac{1}{\zeta_a(0, z)} \\ \frac{\partial}{\partial z} \frac{1}{\zeta} &= \frac{1}{\zeta_s} \frac{\partial}{\partial z} \frac{1}{\zeta_a} + \frac{1}{\zeta_a} \frac{\partial}{\partial z} \frac{1}{\zeta_s}. \end{aligned} \quad (24.19)$$

The leading (material flow conserving) eigenvalue $z = 1$ belongs to the symmetric subspace $1/\zeta_s(0, 1) = 0$, so the derivatives (24.12) also depend only on the symmetric subspace:

$$\begin{aligned} \langle n \rangle_\zeta &= z \frac{\partial}{\partial z} \frac{1}{\zeta(0, z)} \Big|_{z=1} \\ &= \frac{1}{\zeta_a(0, z)} z \frac{\partial}{\partial z} \frac{1}{\zeta_s(0, z)} \Big|_{z=1}. \end{aligned} \quad (24.20)$$

Remark 24.5. Lorentz gas in the fundamental domain. The vector valued nature of the moment-generating function (24.4) in the case under consideration makes it difficult to perform a calculation of the diffusion constant within the fundamental domain. Yet we point out that, at least as regards scalar quantities, the full reduction to $\tilde{\mathcal{M}}$ leads to better estimates. A proper symbolic dynamics in the fundamental domain has been introduced in ref. [12].

In order to perform the full reduction for diffusion one should express the dynamical zeta function (24.8) in terms of the prime cycles of the fundamental domain $\tilde{\mathcal{M}}$ of the lattice (see figure 24.3) rather than those of the elementary (Wigner-Seitz) cell $\mathcal{M}$. This problem is complicated by the breaking of the rotational symmetry by the auxiliary vector β , or, in other words, the non-commutativity of translations and rotations: see ref. [14]. For a 'fundamental domain' in hyperbolic geometry, see for example [these notes](#) by K. Martin.

Remark 24.6. Anomalous diffusion. Anomalous diffusion for 1-dimensional intermittent maps was studied in the continuous time random walk approach in refs. [24, 25].

T_p	# cycles	$\zeta(0,0)$	λ	D
1	5	-0.2169759	1.39193	0.37795
2	10	-0.0248233	1.74541	0.23118
3	33	-0.0221962	1.72235	0.25257
4	108	-0.0002192	1.74450	0.24165
5	373	0.0023463	1.76079	0.24468
6	1378	0.0096330	1.75610	0.24068
numerical experiment			1.760	0.25

Table 24.1: The Lyapunov exponent λ and the diffusion constant D computed in the fundamental domain, $w = 0.3$ disk-disk separation, disks radius = 1 (from Zhanget al. [41]).

The first approach within the framework of cycle expansions (based on truncated dynamical zeta functions) was developed by Artuso *et al.* [2, 4]. For subsequent developments, consult refs. [3, 6] and Klages, Radons and Sokolov [34]. Our treatment follows methods introduced in ref. [15], applied there to investigate the behavior of a Lorentz gas with unbounded horizon.

Question 24.1. Alice wants to know

Q Do these Jonquière functions appear in physics?

A In statistical mechanics Jonquière function (24.35) appears in the theory of free Bose-Einstein gas, see refs. [17, 18].

exercise 29.1

References

- [1] R. Artuso, “Diffusive dynamics and periodic orbits of dynamic systems”, *Phys. Lett. A* **160**, 528–530 (1991).
- [2] R. Artuso, “Recycling deterministic diffusion”, *Physica D* **76**, 1–7 (1994).
- [3] R. Artuso and R. Burioni, “Anomalous diffusion: Deterministic and stochastic perspectives”, in *Large Deviations in Physics: The Legacy of the Law of Large Numbers*, edited by A. Vulpiani, F. Cecconi, M. Cencini, A. Puglisi, and D. Vergni (Springer, Berlin, 2014), pp. 263–293.
- [4] R. Artuso, G. Casati, and R. Lombardi, “Periodic orbit theory of anomalous diffusion”, *Phys. Rev. Lett.* **71**, 62 (1993).
- [5] R. Artuso, G. Casati, and R. Lombardi, “Periodic orbit theory of deterministic diffusion”, *Physica A* **205**, 412–419 (1994).
- [6] R. Artuso and G. Cristadoro, “Deterministic (anomalous) transport”, in *Anomalous Transport: Foundations and Applications*, edited by R. Klages, G. Radons, and I. M. Sokolov (Wiley, New York, 2008).
- [7] P. M. Bleher, “Statistical properties of two-dimensional periodic Lorentz gas with infinite horizon”, *J. Stat. Phys.* **66**, 479–497 (1992).
- [8] L. A. Bunimovich, “Decay of correlations in dynamical systems with chaotic behavior”, *Sov. Phys. JETP* **62**, 842–852 (1985).

- [9] L. A. Bunimovich and Y. G. Sinai, “Markov partitions for dispersed billiards”, *Commun. Math. Phys.* **78**, Erratum, *ibid.* **107**, 357 (1986), 247–280 (1980).
- [10] L. A. Bunimovich and Y. G. Sinai, “Statistical properties of Lorentz gas with periodic configuration of scatterers”, *Commun. Math. Phys.* **78**, 479–497 (1981).
- [11] L. A. Bunimovich, Y. G. Sinai, and N. I. Chernov, “Markov partitions for two-dimensional hyperbolic billiards”, *Russ. Math. Surv.* **45**, 105–152 (1990).
- [12] F. Christiansen, Analysis of Chaotic Dynamical Systems in Terms of Cycles, MA thesis (Univ. of Copenhagen, Copenhagen, 1989).
- [13] F. A. Cotton, *Chemical Applications of Group Theory*, 3rd ed. (Wiley, New York, 2008).
- [14] P. Cvitanović, J.-P. Eckmann, and P. Gaspard, “Transport properties of the Lorentz gas in terms of periodic orbits”, *Chaos Solit. Fract.* **6**, 113–120 (1995).
- [15] P. Cvitanović, P. Gaspard, and T. Schreiber, “Investigation of the Lorentz gas in terms of periodic orbits”, *Chaos* **2**, 85–90 (1992).
- [16] C. P. Dettmann, “Diffusion in the Lorentz gas”, *Commun. Theor. Phys* **62**, 521–540 (2014).
- [17] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions*, Vol. 1, Bateman Manuscript Project (McGraw-Hill, 1953).
- [18] B. Fornberg and K. S. Kölbig, “Complex zeros of the Jonquière or polylogarithm function”, *Math. Comp.* **29**, 582–599 (1975).
- [19] G. Gallavotti, “Lectures on the billiard”, in *Dynamical Systems, Theory and Applications*, Vol. 38, edited by J. Moser, Lecture Notes in Physics (Springer, Berlin, 1975), pp. 236–295.
- [20] G. Gallavotti and E. G. D. Cohen, “Dynamical ensembles in nonequilibrium statistical mechanics”, *Phys. Rev. Lett.* **74**, 2694–2697 (1995).
- [21] P. Gaspard and F. Baras, Dynamical chaos underlying diffusion in the Lorentz gas, in *Microscopic simulations of complex hydrodynamic phenomena*, edited by M. Mareschal and B. L. Holian (1992) Chap. Microscopic simulations of complex hydrodynamic phenomena, pp. 301–322.
- [22] P. Gaspard and G. Nicolis, “Transport properties, Lyapunov exponents, and entropy per unit time”, *Phys. Rev. Lett.* **65**, 1693–1696 (1990).
- [23] T. Geisel and J. Nierwetberg, “Onset of diffusion and universal scaling in chaotic systems”, *Phys. Rev. Lett.* **48**, 7 (1982).
- [24] T. Geisel, J. Nierwetberg, and A. Zacherl, “Accelerated diffusion in Josephson junctions and related chaotic systems”, *Phys. Rev. Lett.* **54**, 616 (1985).
- [25] T. Geisel and S. Thomae, “Anomalous diffusion in intermittent chaotic systems”, *Phys. Rev. Lett.* **52**, 1936 (1984).

- [26] S. Grossmann and H. Fujisaka, “Diffusion in discrete nonlinear dynamical systems”, *Phys. Rev. A* **26**, 1779–1782 (1982).
- [27] S. Grossmann and S. Thomaes, “Shape dependence of correlation times in chaos-induced diffusion”, *Phys. Lett. A* **97**, 263–267 (1983).
- [28] G. Keller, P. J. Howard, and R. Klages, “Continuity properties of transport coefficients in simple maps”, *Nonlinearity* **21**, 1719–1743 (2008).
- [29] R. Klages, *Deterministic Diffusion in One-dimensional Chaotic Dynamical Systems* (Wissenschaft and Technik-Verlag, Berlin, 1996).
- [30] R. Klages, *Microscopic Chaos, Fractals and Transport in Nonequilibrium Statistical Mechanics* (World Scientific, Singapore, 2007).
- [31] R. Klages, “From deterministic chaos to anomalous diffusion”, *Reviews of Nonlinear Dynamics and Complexity*, 169–227 (2010).
- [32] R. Klages and J. R. Dorfman, “Simple maps with fractal diffusion coefficients”, *Phys. Rev. Lett.* **74**, 387–390 (1995).
- [33] R. Klages and J. R. Dorfman, “Dynamical crossover in deterministic diffusion”, *Phys. Rev. E* **55**, R1247–R1250 (1997).
- [34] R. Klages, G. Radons, and I. M. Sokolov, *Anomalous Transport: Foundations and Applications* (Wiley, New York, 2008).
- [35] J. Lloyd, M. Niemeyer, L. Rondoni, and G. P. Morriss, “The nonequilibrium Lorentz gas”, *Chaos* **5**, 536–551 (1995).
- [36] H. A. Lorentz, “The motion of electrons in metallic bodies”, *K. Ned. Akad. van Wet. B* **7**, 438–453 (1905).
- [37] J. Machta and R. Zwanzig, “Diffusion in a periodic Lorentz gas”, *Phys. Rev. Lett.* **50**, 1959–1962 (1983).
- [38] M. Schell, S. Fraser, and R. Kapral, “Diffusive dynamics in systems with translational symmetry: A one-dimensional-map model”, *Phys. Rev. A* **26**, 504–521 (1982).
- [39] Y. G. Sinai, “Dynamical systems with elastic reflections”, *Russ. Math. Surv.* **25**, 137–189 (1970).
- [40] W. N. Vance, “Unstable periodic-orbits and transport-properties of nonequilibrium steady-states”, *Phys. Rev. Lett.* **69**, 1356–1359 (1992).
- [41] T. Zhang, P. Cvitanović, and D. I. Goldman, Diffuse globally, compute locally: a cyclist tale, In preparation, 2017.

24.4 Examples

Example 24.1. Chains of piecewise linear maps. We start by defining the map $\hat{f}$ on the unit interval as

$$\hat{f}(\hat{x}) = \begin{cases} \Lambda \hat{x} & \hat{x} \in [0, 1/2) \\ \Lambda \hat{x} + 1 - \Lambda & \hat{x} \in (1/2, 1] \end{cases}, \quad \Lambda > 2, \quad (24.21)$$

and then extending the dynamics to the entire real line, by imposing the translation property

$$\hat{f}(\hat{x} + \hat{n}) = \hat{f}(\hat{x}) + \hat{n} \quad \hat{n} \in \mathbb{Z}. \quad (24.22)$$

As the map is discontinuous at $\hat{x} = 1/2$, $\hat{f}(1/2)$ is undefined, and the $x = 1/2$ point has to be excluded from the Markov partition. The map is antisymmetric under the $\hat{x}$ -coordinate flip

$$\hat{f}(\hat{x}) = -\hat{f}(-\hat{x}), \quad (24.23)$$

so the dynamics will exhibit no mean drift; all odd derivatives of the moment-generating function (20.10) with respect to β , evaluated at $\beta = 0$, will vanish.

The map (24.21) is sketched in figure 24.4(a). Initial points sufficiently close to either of the fixed points in the initial unit interval remain in the elementary cell for one iteration; depending on the slope Λ , other points jump $\hat{n}$ cells, either to the right or to the left. Repetition of this process generates a random walk for almost every initial condition.

The translational symmetry (24.22) relates the unbounded dynamics on the real line to dynamics restricted to the elementary cell - in the example at hand, the unit interval curled up into a circle. Associated to $\hat{f}(\hat{x})$ we thus also consider the circle map

$$f(x) = \hat{f}(\hat{x}) - [\hat{f}(\hat{x})], \quad x = \hat{x} - [\hat{x}] \in [0, 1] \quad (24.24)$$

figure 24.4(b), where $[\cdots]$ stands for the integer part (24.3). For the piecewise linear map of figure 24.4 we can evaluate the dynamical zeta function in closed form. Each branch has the same value of the slope, and the map can be parameterized by a single parameter, for example its critical value $a = \hat{f}(1/2)$, the absolute maximum on the interval $[0, 1]$ related to the slope of the map by $a = \Lambda/2$. The larger Λ is, the stronger is the stretching action of the map.

[return to p. 480](#)

Example 24.2. Unrestricted symbolic dynamics. Whenever Λ is an integer number, the symbolic dynamics is exceedingly simple. For example, for the case $\Lambda = 6$ illustrated in figure 24.4(b), the elementary cell map consists of 6 full branches, with uniform stretching factor $\Lambda = 6$. The branches have different jumping numbers: for branches 1 and 2 we have $\hat{n} = 0$, for branch 3 we have $\hat{n} = +1$, for branch 4 $\hat{n} = -1$, and finally for branches 5 and 6 we have respectively $\hat{n} = +2$ and $\hat{n} = -2$. The same structure reappears whenever Λ is an even integer $\Lambda = 2a$: all branches are mapped onto the whole unit interval and we have two $\hat{n} = 0$ branches, one branch for which $\hat{n} = +1$ and one for which $\hat{n} = -1$, and so on, up to the maximal jump $|\hat{n}| = a - 1$. The symbolic dynamics is thus full, unrestricted shift in $2a$ symbols $\{0_+, 1_+, \dots, (a-1)_+, (a-1)_-, \dots, 1_-, 0_-\}$, where the symbol indicates both the length and the direction of the corresponding jump.

For the piecewise linear maps with uniform stretching the weight associated with a given symbol sequence is a product of weights for individual steps, $t_{sq} = t_s t_q$. For the map

of figure 24.4 there are 6 distinct weights (24.10):

$$\begin{aligned} t_1 &= t_2 = z/\Lambda \\ t_3 &= e^\beta z/\Lambda, \quad t_4 = e^{-\beta} z/\Lambda, \quad t_5 = e^{2\beta} z/\Lambda, \quad t_6 = e^{-2\beta} z/\Lambda. \end{aligned}$$

The piecewise linearity and the simple symbolic dynamics lead to the full cancelation of all curvature corrections in (23.8). The *exact* dynamical zeta function (18.13) is given by the fixed point contributions:

$$\begin{aligned} 1/\zeta(\beta, z) &= 1 - t_{0+} - t_{0-} - \cdots - t_{(a-1)+} - t_{(a-1)-} \\ &= 1 - \frac{z}{a} \left(1 + \sum_{j=1}^{a-1} \cosh(\beta j) \right). \end{aligned} \quad (24.25)$$

The leading (and only) eigenvalue of the evolution operator (24.6) is

$$s(\beta) = \log \left\{ \frac{1}{a} \left(1 + \sum_{j=1}^{a-1} \cosh(\beta j) \right) \right\}, \quad \Lambda = 2a, \text{ } a \text{ integer}. \quad (24.26)$$

The flow conservation (23.17) sum rule is manifestly satisfied, so $s(0) = 0$. The first derivative $s(0)'$ vanishes as well by the left/right symmetry of the dynamics, implying vanishing mean drift $\langle \hat{x} \rangle = 0$. The second derivative $s(\beta)''$ yields the diffusion constant (24.11):

$$\langle n \rangle_\zeta = 2a \frac{1}{\Lambda} = 1, \quad \langle \hat{x}^2 \rangle_\zeta = 2 \frac{0^2}{\Lambda} + 2 \frac{1^2}{\Lambda} + 2 \frac{2^2}{\Lambda} + \cdots + 2 \frac{(a-1)^2}{\Lambda} \quad (24.27)$$

Using the identity $\sum_{k=1}^n k^2 = n(n+1)(2n+1)/6$ we obtain

$$D = \frac{1}{24}(\Lambda - 1)(\Lambda - 2), \quad \Lambda \text{ even integer}. \quad (24.28)$$

Similar calculation for odd integer $\Lambda = 2k - 1$ yields

$$D = \frac{1}{24}(\Lambda^2 - 1), \quad \Lambda \text{ odd integer}. \quad (24.29)$$

exercise 24.1

return to p. 481

Example 24.3. $\mathcal{B}_4$ Burnett coefficient. For the map under consideration the first Burnett coefficient coefficient $\mathcal{B}_4$ (or kurtosis (A20.11)) is easily evaluated. For example, using (24.26) in the case of even integer slope $\Lambda = 2a$ we obtain

exercise 24.2

$$\mathcal{B}_4 = -\frac{1}{4! \cdot 60} (a-1)(2a-1)(4a^2 - 9a + 7). \quad (24.30)$$

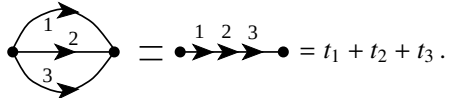
return to p. 482

Example 24.4. A finite Markov partition. As an example we determine a value of the parameter $4 \leq \Lambda \leq 6$ for which $f(f(1/2)) = 0$. As in the integer Λ case, we partition the unit interval into six intervals, labeled by the jumping number $\hat{n}(x) \in \{\mathcal{M}_{0+}, \mathcal{M}_{1+}, \mathcal{M}_{2+}, \mathcal{M}_{2-}, \mathcal{M}_{1-}, \mathcal{M}_{0-}\}$, ordered by their placement along the unit interval, figure 24.5 (a).

In general the critical value $a = \hat{f}(1/2)$ will not correspond to an interval border, but now we choose a such that the critical point is mapped onto the right border of $\mathcal{M}_{1+}$. Equating $f(1/2)$ with the right border of $\mathcal{M}_{1+}$, $x = 1/\Lambda$, we obtain a quadratic equation with the expanding solution $\Lambda = 2(\sqrt{2} + 1)$. For this parameter value $f(\mathcal{M}_{2+}) = \mathcal{M}_{0+} \cup \mathcal{M}_{1+}$, $f(\mathcal{M}_{2-}) = \mathcal{M}_{0-} \cup \mathcal{M}_{1-}$, while the remaining intervals map onto the whole unit interval $\mathcal{M}$. The transition matrix (17.1) is given by

$$\phi' = T\phi = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \phi_{0+} \\ \phi_{1+} \\ \phi_{2+} \\ \phi_{2-} \\ \phi_{1-} \\ \phi_{0-} \end{bmatrix}. \quad (24.31)$$

One could diagonalize (24.31) on a computer, but, as we saw in chapter 17, the transition graph of figure 24.5 (b) corresponding to map figure 24.5 (a) offers more insight into the dynamics. Figure 24.5 (b) can be redrawn more compactly as transition graph figure 24.5 (c) by replacing parallel lines in a graph by their sum



$$= \bullet \xrightarrow{1+2+3} \bullet = t_1 + t_2 + t_3. \quad (24.32)$$

The dynamics is unrestricted in the alphabet

$$\mathcal{A} = \{0_+, 1_+, 2_+0_+, 2_+1_+, 2_-1_-, 2_-0_-, 1_-, 0_-\}.$$

Applying the loop expansion (18.13) of sect. 18.3, we are led to the dynamical zeta function

$$\begin{aligned} 1/\zeta(\beta, z) &= 1 - t_{0+} - t_{1+} - t_{2+0+} - t_{2+1+} - t_{2-1-} - t_{2-0-} - t_{1-} - t_{0-} \\ &= 1 - \frac{2z}{\Lambda} (1 + \cosh(\beta)) - \frac{2z^2}{\Lambda^2} (\cosh(2\beta) + \cosh(3\beta)). \end{aligned} \quad (24.33)$$

For grammar as simple as this one, the dynamical zeta function is the sum over fixed points of the unrestricted alphabet. As the first check of this expression for the dynamical zeta function we verify that

$$1/\zeta(0, 1) = 1 - \frac{4}{\Lambda} - \frac{4}{\Lambda^2} = 0,$$

as required by the flow conservation (23.17). Conversely, we could have started by picking the desired Markov partition, writing down the corresponding dynamical zeta function, and then fixing Λ by the $1/\zeta(0, 1) = 0$ condition. For more complicated transition graphs this approach, together with the factorization (24.19), is helpful in reducing the order of the polynomial condition that fixes Λ .

The diffusion constant follows from (24.11)

exercise 24.3

$$\begin{aligned} \langle n \rangle_\zeta &= 4 \frac{1}{\Lambda} + 4 \frac{2}{\Lambda^2}, \quad \langle \hat{n}^2 \rangle_\zeta = 2 \frac{1^2}{\Lambda} + 2 \frac{2^2}{\Lambda^2} + 2 \frac{3^2}{\Lambda^2} \\ D &= \frac{15 + 2\sqrt{2}}{16 + 8\sqrt{2}}. \end{aligned} \quad (24.34)$$

return to p. 482

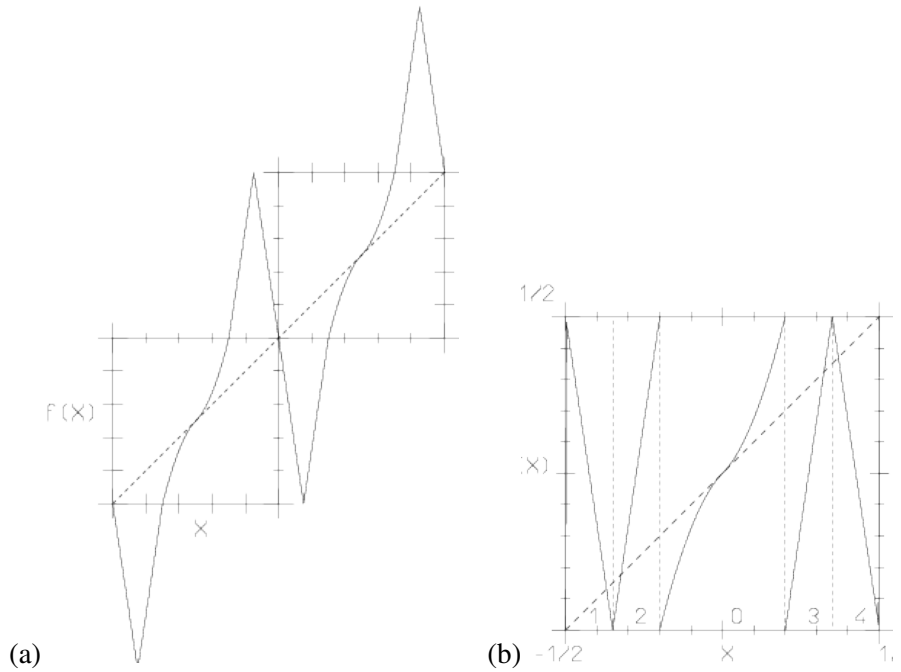


Figure 24.7: (a) A map with marginal fixed point.
(b) The map restricted to the unit circle.

Example 24.5. Anomalous diffusion. Consider a 1-dimensional map of the real line on itself shown in figure 24.7 (a), with the same properties as in sect. 24.2, except for a marginal fixed point at $x = 0$. The corresponding circle map is given in figure 24.7 (b).

As in sect. 29.2.1, a branch with support in M_i , $i = 1, 2, 3, 4$ has constant slope Λ_i , while $f|_{M_0}$ is of intermittent form. To keep you nimble, this time we take a slightly different choice of slopes. The toy example of sect. 29.2.1 was cooked up so that the $1/s$ branch cut in dynamical zeta function was the whole answer. Here we shall take a slightly different route, and pick piecewise constant slopes such that the dynamical zeta function for intermittent system can be expressed in terms of the Jonquière function

question 24.1

$$J(z, s) = \sum_{k=1}^{\infty} z^k / k^s. \quad (24.35)$$

Once the $\bar{0}$ fixed point is pruned away, the symbolic dynamics is given by the infinite alphabet $\{1, 2, 3, 4, 0^l 1, 0^l 2, 0^k 3, 0^l 4\}$, $i, j, k, l = 1, 2, \dots$ (compare with table 29.1). The partitioning of the subinterval M_0 is induced by $M_{0^k(\text{right})} = \hat{f}_{(\text{right})}^{-k}(M_3 \cup M_4)$ (where $\hat{f}_{(\text{right})}^{-1}$ denotes the inverse of the right branch of $\hat{f}|_{M_0}$) and the same reasoning applies to the leftmost branch. These are regions over which the slope of $\hat{f}|_{M_0}$ is constant. Thus we have the following stabilities and jumping numbers associated to letters:

$$\begin{array}{lll} 0^k 3, 0^k 4 & \Lambda_p = \frac{k^{1+\alpha}}{q/2} & \hat{n}_p = 1 \\ 0^l 1, 0^l 2 & \Lambda_p = \frac{l^{1+\alpha}}{q/2} & \hat{n}_p = -1 \\ 3, 4 & \Lambda_p = \pm \Lambda & \hat{n}_p = 1 \\ 2, 1 & \Lambda_p = \pm \Lambda & \hat{n}_p = -1, \end{array} \quad (24.36)$$

where $\alpha = 1/s$ is determined by the intermittency exponent (29.1), while q is to be determined by the flow conservation (23.17) for $\hat{f}$:

$$\frac{4}{\Lambda} + 2q\zeta(\alpha + 1) = 1$$

(where ζ is the Riemann zeta function), so that $q = (\Lambda - 4)/(2\Lambda\zeta(\alpha + 1))$. The dynamical zeta function picks up contributions just by the alphabet's letters, as we have imposed piecewise linearity, and can be expressed in terms of a Jonquière function (24.35):

$$1/\zeta_0(z, \beta) = 1 - \frac{4}{\Lambda} z \cosh \beta - \frac{\Lambda - 4}{\Lambda \zeta(1 + \alpha)} z \cosh \beta \cdot J(z, \alpha + 1). \quad (24.37)$$

Its first zero $z(\beta)$ is determined by

$$\frac{4}{\Lambda} z + \frac{\Lambda - 4}{\Lambda \zeta(1 + \alpha)} z \cdot J(z, \alpha + 1) = \frac{1}{\cosh \beta}.$$

[return to p. 483](#)

24.5 Exercises

24.1. **Diffusion for odd integer Λ .** Show that when the slope $\Lambda = 2k - 1$ in (24.21) is an odd integer, the diffusion constant is given by $D = (\Lambda^2 - 1)/24$, as stated in (24.29).

24.2. **Fourth-order transport coefficient.** Verify (24.30). You will need the identity

$$\sum_{k=1}^n k^4 = \frac{1}{30} n(n+1)(2n+1)(3n^2 + 3n - 1).$$

24.3. **Finite Markov partitions.** Verify (24.34).

24.4. **Maps with variable peak shape:** Consider the following piecewise linear map

$$f_\delta(x) = \begin{cases} \frac{3x}{1-\delta} & x \in \mathcal{M}_1 \\ \frac{3}{2} - \left(\frac{2}{\delta} \left| \frac{4-\delta}{12} - x \right| \right) & x \in \mathcal{M}_2 \\ 1 - \frac{3}{1-\delta} \left(x - \frac{1}{6}(2+\delta) \right) & x \in \mathcal{M}_3 \end{cases}$$

where $\mathcal{M}_1 = [0, \frac{1}{3}(1-\delta)]$, $\mathcal{M}_2 = [\frac{1}{3}(1-\delta), \frac{1}{6}(2+\delta)]$, $\mathcal{M}_3 = [\frac{1}{6}(2+\delta), \frac{1}{2}]$, and the map in $[1/2, 1]$ is obtained by antisymmetry with respect to $x = 1/2, y = 1/2$. Write the corresponding dynamical zeta function relevant to diffusion and then show that

$$D = \frac{\delta(2+\delta)}{4(1-\delta)}$$

See refs. [5, 27] for further details.

24.5. **Two-symbol cycles for the Lorentz gas.** Write down all cycles labeled by two symbols, such as (0 6), (1 7), (1 5) and (0 5).

ChaosBook.org/projects offers several project-length deterministic diffusion exercises.

24.6. **Accelerated diffusion.** (medium difficulty) Consider a map h , such that $\hat{h} = \hat{f}$ of figure 24.7 (b), but now running branches are turned into standing branches and vice versa, so that 1, 2, 3, 4 are standing while 0 leads to both positive and negative jumps. Build the corresponding dynamical zeta function and show that

$$\sigma^2(t) \sim \begin{cases} t & \text{for } \alpha > 2 \\ t \ln t & \text{for } \alpha = 2 \\ t^{3-\alpha} & \text{for } \alpha \in (1, 2) \\ t^2 / \ln t & \text{for } \alpha = 1 \\ t^2 & \text{for } \alpha \in (0, 1) \end{cases}$$

24.7. **Recurrence times for Lorentz gas with infinite horizon.** Consider the Lorentz gas with unbounded horizon with a square lattice geometry, with disk radius R and unit lattice spacing. Label disks according to the (integer) coordinates of their center: the sequence of recurrence times $\{t_j\}$ is given by the set of collision times. Consider orbits that leave the disk sitting at the origin and hit a disk far away after a free flight (along the horizontal corridor). Initial conditions are characterized by coordinates (ϕ, α) (ϕ determines the initial position along the disk, while α gives the angle of the initial velocity with respect to the outward normal: the appropriate measure is then $d\phi \cos \alpha d\alpha$ ($\phi \in [0, 2\pi)$, $\alpha \in [-\pi/2, \pi/2]$). Find how $\phi(T)$ scales for large values of T : this is equivalent to investigating the scaling of portions of the state space that lead to a first collision with disk $(n, 1)$, for large values of n (as $n \mapsto \infty$ $n \simeq T$).

24.8. Diffusion reduced to the fundamental domain.

Maps such as figure 24.4 are antisymmetric. Reduce such antisymmetric maps as in example 10.5, and write down the formula (24.11) for the diffusion constant D in terms of the fundamental domain cycles (relative periodic orbits) alone (P. Gaspard says it cannot be done [14]).

Chapter 25

Discrete symmetry factorization

No endeavor that is worthwhile is simple in prospect; if it is right, it will be simple in retrospect.

—Edward Teller

TO THOSE VERSED in Quantum Mechanics (QM), utility of symmetries in reducing spectrum calculations is *sine qua non*: if a group of symmetries commutes with the Hamiltonian, irreducible representations of the symmetry group block-diagonalize it, each block spanned by a set of the degenerate eigenstates of the same energy. Like most QM gymnastics, this block-diagonalization has nothing to do with quantum mysteries, it is just linear algebra. As we shall show here, deterministic spectral determinants factor in the same way, given that the evolution operator $\mathcal{L}^t(y, x)$ for a system $f^t(x)$ is invariant under a discrete symmetry group $G = \{e, g_2, g_3, \dots, g_{|G|}\}$ of order $|G|$. In the process we 1.) learn that the deterministic dynamics, once recast into the language of evolution operators, is much closer to quantum mechanics than is apparent in the Newtonian, ODE formulation (linear evolution operators, group-theoretical spectral decompositions, ...), 2.) that once the symmetry group is quotiented out, the dynamics simplifies, and 3.) it's a triple home run: simpler symbolic dynamics, fewer cycles needed, much better convergence of cycle expansions. Once you master this, going back to your pre-desymmetrization ways is unthinkable.

The main result of this chapter can be stated as follows:

If the dynamics possesses a discrete symmetry, the contribution of a cycle p of multiplicity m_p (10.6) to a dynamical zeta function factorizes into a product over the d_μ -dimensional irreps $D^{(\mu)}(g)$ of the symmetry group,

$$(1 - t_p)^{m_p} = \prod_{\mu} \det \left(1 - D^{(\mu)}(h_{\hat{p}}) t_{\hat{p}} \right)^{d_{\mu}}, \quad t_p = t_{\hat{p}}^{|G|/m_p},$$

where $t_{\hat{p}}$ is the cycle weight evaluated on the relative periodic orbit $\hat{p}$, $|G|$ is the order of the group, $h_{\hat{p}}$ is the group element relating the fundamental domain cycle $\hat{p}$ to a segment of the full space cycle p , and m_p is the multiplicity of the p cycle.

As dynamical zeta functions have particularly simple cycle expansions, a geometrical shadowing interpretation of their convergence, and suffice for determination of leading eigenvalues, we shall use them to explain the group-theoretic factorizations; the full spectral determinants can be factorized using the same techniques.

This chapter is meant to serve as a detailed guide to the computation of dynamical zeta functions and spectral determinants for systems with discrete symmetries. Familiarity with basic group-theoretic notions is assumed, with some details relegated to appendix A10.1. We develop here the cycle expansions for factorized determinants, and exemplify them by working out two cases of physical interest: D_1 and D_3 symmetries. $D_2 = D_1 \times D_1$ and D_4 symmetries are discussed in appendix A25. We start with a review of some basic facts of the group representation theory.

25.1 Transformation of functions

So far we have recast the problem of long time dynamics into language of linear operators acting on functions, simplest one of which is $\rho(x, t)$, the density of trajectories at time t . First we will explain what discrete symmetries do to such functions, and then how they affect their evolution in time. 

Let g be an *abstract group element* in G . For a discrete group a group element is typically indexed by a discrete label, $g = g_j$. For a continuous group it is typically parametrized by a set of continuous parameters, $g = g(\theta_m)$. As discussed on page 180, linear action of a group element $g \in G$ on a state $x \in \mathcal{M}$ is given by its *matrix representation*, a finite non-singular $[d \times d]$ matrix $D(g)$:

$$x \rightarrow x' = D(g)x. \quad (25.1)$$



example 25.1
p. 517



example 25.2
p. 517

How does the group act on a function ρ of x ? Denote by $U(g)$ the operator $\rho'(x) = U(g)\rho(x)$ that returns the transformed function. One *defines* the transformed function ρ' by requiring that it has the same value at $x' = D(g)x$ as the initial function has at x ,

$$\rho'(x') = U(g)\rho(D(g)x) = \rho(x).$$

Replacing $x \rightarrow D(g)^{-1}x$, we find that a group element $g \in G$ acts on a function $\rho(x)$ defined on state space $\mathcal{M}$ by its *operator representation*

$$U(g)\rho(x) = \rho(D(g)^{-1}x). \quad (25.2)$$

This is the conventional, Wigner definition of the effect of transformations on functions that should be familiar to master quantum mechanics. Again: $U(g)$ is an ‘*operator*’, not a matrix - it is an operation whose only meaning is exactly what

(25.2) says. And yes, Mathilde, the action on the state space points is $D(g)^{-1}x$, not $D(g)x$.

Consider next the effect of two successive transformations g_1, g_2 :

$$\begin{aligned} U(g_2)U(g_1)\rho(x) &= U(g_2)\rho(D(g_1)^{-1}x) = \rho(D(g_2)^{-1}D(g_1)^{-1}x) \\ &= \rho(D(g_1g_2)^{-1}x) = U(g)\rho(x). \end{aligned}$$

Hence if $g_1g_2 = g$, we have $U(g_2)U(g_1) = U(g)$: so operators $U(g)$ form a representation of the group.

25.2 Taking care of fundamentals

Instant gratification takes too long.

— Carrie Fisher

If a dynamical system $(\mathcal{M}, f)$ is equivariant under a discrete symmetry (visualize the 3-disk billiard, figure 10.1), the state space $\mathcal{M}$ can be tiled by a *fundamental domain* $\hat{\mathcal{M}}$ and its images $\hat{\mathcal{M}}_2 = g_2\hat{\mathcal{M}}$, $\hat{\mathcal{M}}_3 = g_3\hat{\mathcal{M}}$, $\dots$ under the action of the symmetry group $G = \{e, g_2, \dots, g_{|G|}\}$,

$$\mathcal{M} = \sum_{g \in G} \hat{\mathcal{M}}_g = \hat{\mathcal{M}} \cup \hat{\mathcal{M}}_2 \cup \hat{\mathcal{M}}_3 \cdots \cup \hat{\mathcal{M}}_{|G|}. \quad (25.3)$$



section 11.3



example 25.3
p. 517

25.2.1 Regular representation

Take an arbitrary function $\rho(x)$ defined over the state space $x \in \mathcal{M}$. If the state space is tiled by a fundamental domain $\hat{\mathcal{M}}$ and its copies, function $\rho(x)$ can be written as a $|G|$ -dimensional vector of functions, each function defined over the fundamental domain $\hat{x} \in \hat{\mathcal{M}}$ only. The natural choice of a function space basis is the $|G|$ -component *regular basis* vector

$$\begin{bmatrix} \rho_1^{reg}(\hat{x}) \\ \rho_2^{reg}(\hat{x}) \\ \vdots \\ \rho_{|G|}^{reg}(\hat{x}) \end{bmatrix} = \begin{bmatrix} \rho(D(e)\hat{x}) \\ \rho(D(g_2)\hat{x}) \\ \vdots \\ \rho(D(g_{|G|})\hat{x}) \end{bmatrix}, \quad (25.4)$$

constructed from an arbitrary function $\rho(x)$ defined over the entire state space $\mathcal{M}$, by applying $U(g^{-1})$ to $\rho(\hat{x})$ for each $g \in G$, with state space points restricted to the fundamental domain, $\hat{x} \in \hat{\mathcal{M}}$.

Now apply group action *operator* $U(g)$ to a regular basis vector:

$$U(g) \begin{bmatrix} \rho(D(e)\hat{x}) \\ \rho(D(g_2)\hat{x}) \\ \vdots \\ \rho(D(g_{|G|})\hat{x}) \end{bmatrix} = \begin{bmatrix} \rho(D(g^{-1})\hat{x}) \\ \rho(D(g^{-1}g_2)\hat{x}) \\ \vdots \\ \rho(D(g^{-1}g_{|G|})\hat{x}) \end{bmatrix}.$$

It acts by permuting the components. (And yes, Mathilde, the pesky g^{-1} is inherited from (25.2), and there is nothing you can do about it.) Thus the action of the *operator* $U(g)$ on a regular basis vector can be represented by the corresponding $[|G| \times |G|]$ permutation *matrix*, called the *left regular representation* $D^{reg}(g)$,

$$U(g) \begin{bmatrix} \rho_1^{reg}(\hat{x}) \\ \rho_2^{reg}(\hat{x}) \\ \vdots \\ \rho_{|G|}^{reg}(\hat{x}) \end{bmatrix} = D^{reg}(g) \begin{bmatrix} \rho_1^{reg}(\hat{x}) \\ \rho_2^{reg}(\hat{x}) \\ \vdots \\ \rho_{|G|}^{reg}(\hat{x}) \end{bmatrix}.$$

A product of two permutations is a permutation, so this is a matrix representation of the group. To compute its entries, write out the matrix multiplication explicitly, labeling the vector components by the corresponding group elements,

$$\rho_b^{reg}(\hat{x}) = \sum_a^G D^{reg}(g)_{ba} \rho_a^{reg}(\hat{x}).$$

A product of two group elements $g^{-1}a$ is a unique element b , so the a_{th} row of $D^{reg}(g)$ is all zeros, except the b_{th} column which satisfies $g = b^{-1}a$. We arrange the columns of the multiplication table by the inverse group elements, as in table 25.1. Setting multiplication table entries with g to 1, and the rest to 0 then defines the regular representation *matrix* $D^{reg}(g)$ for a given g ,

$$D^{reg}(g)_{ab} = \delta_{g,b^{-1}a}. \quad (25.5)$$

For instance, in the case of the 2-element group $\{e, \sigma\}$ the $D^{reg}(g)$ can be either the identity or the interchange of the two domain labels,

$$D^{reg}(e) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad D^{reg}(\sigma) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \quad (25.6)$$

The multiplication table for D_3 is a more typical, nonabelian group example: see table 25.1. The multiplication tables for C_2 and C_3 are given in table 25.2.

The regular representation of group identity element e is always the identity matrix. As $D^{reg}(g)$ is a permutation matrix, mapping a tile $\hat{M}_a$ into a different tile $\hat{M}_{ga} \neq \hat{M}_a$ if $g \neq e$, only $D^{reg}(e)$ has diagonal elements, and

$$\text{tr } D^{reg}(g) = |G| \delta_{g,e}. \quad (25.7)$$



example 25.4
p. 517



example 25.5
p. 518

D_3	e	σ_{12}	σ_{23}	σ_{31}	d	d_2
e	e	σ_{12}	σ_{23}	σ_{31}	d	d_2
σ_{12}^{-1}	σ_{12}	e	d	d_2	σ_{23}	σ_{31}
σ_{23}^{-1}	σ_{23}	d_2	e	d	σ_{31}	σ_{12}
σ_{31}^{-1}	σ_{31}	d	d_2	e	σ_{12}	σ_{23}
d^{-1}	d_2	σ_{23}	σ_{31}	σ_{12}	e	d
d^{-2}	d	σ_{31}	σ_{12}	σ_{23}	d_2	e

$$D^{reg}(\sigma_{23}) = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad D^{reg}(d) = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Table 25.1: (top) The multiplication table of D_3 , the group of symmetries of a triangle. (bottom) By (25.5), the 6 regular representation matrices $D^{reg}(g)$ of dihedral group D_3 have '1' at the location of g in the D_3 multiplication table table 25.1, '0' elsewhere. For example, the regular representation of the action of operators $U(\sigma_{23})$ and $U(d_2)$ on the regular basis (25.4) are shown here.

25.2.2 Irreps: to get invariants, average

A representation $D^{(\mu)}(g)$ acting on d_μ -dimensional vector space $V^{(\mu)}$ is an *irreducible representation (irrep)* of group G if its only invariant subspaces are $V^{(\mu)}$ and the null vector $\{0\}$. To develop a feeling for this, one can train on a number of simple examples, and work out in each case explicitly a similarity transformation S that brings $D^{reg}(g)$ to a block diagonal form

$$S^{-1}D^{reg}(g)S = \begin{bmatrix} D^{(1)}(g) & & \\ & D^{(2)}(g) & \\ & & \ddots \end{bmatrix} \quad (25.8)$$

for every group element g , such that the corresponding subspace is invariant under actions $g \in G$, and contains no further nontrivial subspace within it. For the problem at hand we do not need to construct invariant subspaces $\rho^{(\mu)}(x)$ and $D^{(\mu)}(g)$ explicitly. We are interested in the symmetry reduction of the trace formula, and for that we will need only one simple result (lemma, theorem, whatever): the regular representation of a finite group contains all of its irreps μ , and its trace is given by the sum

$$\text{tr } D^{reg}(g) = \sum_{\mu} d_{\mu} \chi^{(\mu)}(g), \quad (25.9)$$

where d_{μ} is the dimension of irrep μ , and the characters $\chi^{(\mu)}(g)$ are numbers *intrinsic* to the group G that have to be tabulated only once in the history of humanity. And they all have been. The finiteness of the number of irreps and their dimensions d_{μ} follows from the dimension sum rule for $\text{tr } D^{reg}(e)$, $|G| = \sum d_{\mu}^2$.

The simplest example is afforded by the 1-dimensional subspace (irrep) given by the fully symmetrized average of components of the regular basis function

$\rho^{reg}(x)$

$$\rho^{(A_1)}(x) = \frac{1}{|G|} \sum_g^G \rho(D(g)x).$$

By construction, $\rho^{(A_1)}$ is invariant under all actions of the group, $U(g)\rho^{(A_1)}(x) = \rho^{(A_1)}(x)$. In other words, for every g this is an eigenvector of the regular representation $D^{reg}(g)$ with eigenvalue 1. Other eigenvalues, eigenvectors follow by working out C_3 , C_N (discrete Fourier transform!) and D_3 examples.



example 25.6
p. 518



example 25.7
p. 518



example 25.8
p. 519

The beautiful Frobenius ‘character orthogonality’ theory of irreps (irreducible representations) of finite groups follows, and is sketched here in appendix A25; it says that all other invariant subspaces are obtained by weighted averages (‘projections’)



$$\rho^{(\mu)}(x) = \frac{d_\mu}{|G|} \sum_g \chi^{(\mu)}(g) U(g)\rho(x) = \frac{d_\mu}{|G|} \sum_g \chi^{(\mu)}(g) \rho(D(g^{-1})x) \quad (25.10)$$

The above $\rho^{(A_1)}(x)$ invariant subspace is a special case, with all $\chi^{(A_1)}(g) = 1$.

By now the group acts in many different ways, so let us recapitulate:

g	abstract group element, multiplies other elements
$D(g)$	$[d \times d]$ state space transformation matrix, multiplies $x \in \mathcal{M}$
$U(g)$	operator, acts on functions $\rho(x)$ defined over state space $\mathcal{M}$
$D^{(\mu)}(g)$	$[d_\mu \times d_\mu]$ irrep, acts on invariant subspace $x \in \mathcal{M}^{(\mu)}$
$D^{reg}(g)$	$[G \times G]$ regular matrix rep, acts on vectors $x \in \mathcal{M}^{reg}$

Note that the state space transformation $D(g) \neq D(e)$ can leave sets of ‘boundary’ points invariant (or ‘invariant points’, see (10.10)); for example, under reflection σ across a symmetry plane, the plane itself remains invariant. The boundary periodic orbits that belong to such pointwise invariant sets will require special care in evaluations of trace formulas.

25.3 Dynamics in the fundamental domain

What happens in the fundamental domain, stays in the fundamental domain.

—Prof. Dr. Dr. h.c. Gatto Nero, Grande Ufficiale dell’Ordine della Stella della Notorietà Italiana

How does a group act on the evolution operator $\mathcal{L}^t(y, x)$? As in (25.2), its value should be the same if evaluated at the same points in the rotated coordinates,

$$U(g)\mathcal{L}^t(y, x) = \mathcal{L}^t(D(g)^{-1}y, D(g)^{-1}x). \quad (25.11)$$

We are interested in a dynamical system invariant under the symmetry group G , i.e., with equations of motion invariant (equivariant) under all symmetries $g \in G$, section 10.1

$$D(g) f^t(x) = f^t(D(g) x), \quad (25.12)$$

hence for the evolution operator defined by (20.14) (we can omit the observable weight with no loss of generality, as long as the observable does not break the symmetry):

$$\begin{aligned} U(g^{-1}) \mathcal{L}^t(y, x) &= \mathcal{L}^t(D(g) y, D(g) x) \\ &= \delta(D(g) y - f^t(D(g) x)) = \delta(D(g) (y - f^t(x))) \\ &= \frac{1}{|\det D(g)|} \delta(y - f^t(x)). \end{aligned}$$

For compact groups $|\det D(g)| = 1$ by (10.3), so the evolution operator $\mathcal{L}^t(y, x)$ is *invariant* under group actions,

$$U(g) \mathcal{L}^t(y, x) = \mathcal{L}^t(y, x). \quad (25.13)$$

This is as it should be. If G is a symmetry of dynamics, the law that moves densities around should have the same form in all symmetry related coordinate systems.

As the function $\rho(x)$ that the evolution operator (20.14) acts on is now replaced by the regular basis vector of functions (25.4) over the fundamental domain, the evolution operator itself becomes a $[|G| \times |G|]$ matrix. If the initial point lies in tile $\hat{\mathcal{M}}_a$, its deterministic trajectory lands in the unique tile $\hat{\mathcal{M}}_b$, with a unique relative shift $g = b^{-1}a$, with the only non-vanishing entry $\mathcal{L}^t(y, x)_{ba} = \mathcal{L}^t(D(b)\hat{y}, D(a)\hat{x})$ wherever the regular representation $D^{reg}(g)_{ba}$ has entry 1 in row a and column b . Using the evolution operator invariance (25.13) one can move the end point y into the fundamental domain, and then use the relation $g = b^{-1}a$ to relate the start point x to its image in the fundamental domain,  

$$\mathcal{L}^t(D(b)\hat{y}, D(a)\hat{x}) = \mathcal{L}^t(\hat{y}, D(g)\hat{x}) \equiv \hat{\mathcal{L}}^t(\hat{y}, \hat{x}; g).$$

For a given g all non-vanishing entries are the same, and the evolution operator (20.14) is replaced by the $[|G| \times |G|]$ matrix of form

$$\mathcal{L}_{ba}^t(\hat{y}, \hat{x}; g) = D^{reg}(g)_{ba} \hat{\mathcal{L}}^t(\hat{y}, \hat{x}; g),$$

if $\hat{x} \in \hat{\mathcal{M}}_a$ and $\hat{y} \in \hat{\mathcal{M}}_b$, zero otherwise, and the evolution $\hat{\mathcal{L}}^t(\hat{y}, \hat{x}; g)$ restricted to $\hat{\mathcal{M}}$. Another way to say it is that the law of evolution in the fundamental domain is given by 

$$\hat{x}(t) = \hat{f}^t(\hat{x}_0) = D(g(t)) f^t(\hat{x}_0),$$

where the matrix $D(g(t))$ is the group operation that maps the end point of the full state space trajectory $x(t)$ back to its fundamental domain representative $\hat{x}(t)$. While the global trajectory runs over the full space $\mathcal{M}$, the symmetry-reduced trajectory is brought back into the fundamental domain $\hat{\mathcal{M}}$ every time it crosses

into an adjoining tile; the two trajectories are related by the ‘reconstruction’ operation $g = g(\hat{x}_0, t)$ which maps the global trajectory endpoint into its fundamental domain image.



Now the traces (22.3) required for the evaluation of the eigenvalues of the evolution operator can be computed on the fundamental domain alone

$$\mathrm{tr} \mathcal{L}^t = \int_{\mathcal{M}} dx \mathcal{L}^t(x, x) = \sum_g^G \mathrm{tr} D^{reg}(g) \int_{\hat{\mathcal{M}}} d\hat{x} \hat{\mathcal{L}}^t(\hat{x}, \hat{x}; g). \quad (25.14)$$

Nothing seems to have been gained: the trace of regular representation matrix $\mathrm{tr} D^{reg}(g) = |G| \delta_{g,e}$ guarantees that only those repeats of the fundamental domain cycles $\hat{p}$ that correspond to complete global cycles p contribute, and the factor $\mathrm{tr} D^{reg}(e) = |G|$ simply says that integral over whole state space is $|G|$ times the integral over the fundamental domain.



example 25.10
p. 521

But not so fast! Nobody said that the traces of the *irreps*, $\mathrm{tr} D^{(\mu)}(g) = \chi^{(\mu)}(g)$, in the decomposition (25.9) are nonvanishing only for the identity operation e ; they pick up a contribution for every reconstruction operation $g(\hat{x}_0, t)$,



$$\mathrm{tr} \mathcal{L}^t = \sum_{\mu} d_{\mu} \mathrm{tr} \hat{\mathcal{L}}_{\mu}^t, \quad \mathrm{tr} \hat{\mathcal{L}}_{\mu}^t = \sum_g^G \chi^{(\mu)}(g) \int_{\hat{\mathcal{M}}} d\hat{x} \hat{\mathcal{L}}^t(\hat{x}, \hat{x}; g), \quad (25.15)$$

and then the fundamental domain trace $\int d\hat{x} \hat{\mathcal{L}}^t(\hat{x}, \hat{x}; g)$ picks up a contribution from each fundamental domain prime cycle $\hat{p}$, i.e., all *relative periodic orbits*



$$\hat{x}_{\hat{p}} = g_{\hat{p}} f^{T_{\hat{p}}}(\hat{x}_{\hat{p}}), \quad g_{\hat{p}} = g(\hat{x}_{\hat{p}}, T_{\hat{p}}).$$

In chapter 11 we have shown that a discrete symmetry induces degeneracies among periodic orbits and decomposes periodic orbits into repetitions of irreducible segments; this reduction to a fundamental domain furthermore leads to a convenient symbolic dynamics compatible with the symmetry, and, most importantly, to a factorization of dynamical zeta functions. This we now develop, first in a general setting and then for specific examples.

25.4 Discrete symmetry factorizations

As we saw in chapter 11, discrete symmetries relate classes of periodic orbits and reduce dynamics to a fundamental domain. Such symmetries simplify and improve the cycle expansions in a rather beautiful way; in deterministic dynamics, just as in quantum mechanics, the symmetrized subspaces can be probed by linear operators of different symmetries. If a linear operator commutes with the symmetry, it can be block-diagonalized, and, as we shall now show, the associated spectral determinants and dynamical zeta functions factorize.

We start by working out the factorization of dynamical zeta functions for reflection-symmetric systems in sect. 25.5, and the factorization of the corresponding spectral determinants in example 25.9. As reflection symmetry is essentially the only discrete symmetry that a map of the interval can have, this example completes the group-theoretic factorization of determinants and zeta functions for 1-dimensional maps.

25.4.1 Factorization of dynamical zeta functions

Let p be the full orbit, $\hat{p}$ the orbit in the fundamental domain and $h_{\hat{p}}$ an element of G_p , the symmetry group of p . Restricting the volume integrations to the infinitesimal neighborhoods of the cycles p and $\hat{p}$, respectively, and performing the standard resummations, we obtain the identity

$$(1 - t_p)^{m_p} = \det \left(1 - D^{reg}(h_{\hat{p}})t_{\hat{p}} \right), \quad (25.16)$$

valid cycle by cycle in the Euler products (22.11) for the dynamical zeta function. Here ‘det’ refers to the $[|G| \times |G|]$ regular matrix representation $D^{reg}(h_{\hat{p}})$; as we shall see, this determinant can be evaluated in terms of irrep characters, and no explicit representation of $D^{reg}(h_{\hat{p}})$ is needed. Finally, if a cycle p is invariant under the symmetry subgroup $G_p \subseteq G$ of order h_p , its weight can be written as a repetition of a fundamental domain cycle

$$t_p = t_{\hat{p}}^{h_p} \quad (25.17)$$

computed on the irreducible segment that corresponds to a fundamental domain cycle.

According to (25.16) and (25.17), the contribution of a degenerate class of global cycles (cycle p with multiplicity (10.6) $m_p = |G|/|G_p|$) to a dynamical zeta function is given by the corresponding fundamental domain cycle $\hat{p}$:

$$(1 - t_{\hat{p}}^{|G_p|})^{m_p} = \det \left(1 - D^{reg}(g_{\hat{p}})t_{\hat{p}} \right) \quad (25.18)$$

Let $D^{reg}(g) = \bigoplus_{\mu} d_{\mu} D^{(\mu)}(g)$ be the decomposition of the regular matrix representation into the d_{μ} -dimensional irreps μ of a finite group G . Such decompositions are block-diagonal, so the corresponding contribution to the Euler product (22.8) factorizes as

$$\det(1 - D^{reg}(g)t) = \prod_{\mu} \det(1 - D^{(\mu)}(g)t)^{d_{\mu}}, \quad (25.19)$$

where now the product extends over all distinct d_{μ} -dimensional irreps, each contributing d_{μ} times. For the cycle expansion purposes, it has been convenient to emphasize that the group-theoretic factorization can be effected cycle by cycle, as in (25.18); but from the evolution operator point of view, the key observation is that the symmetry reduces the evolution operator to a block diagonal form; this

block diagonalization implies that the dynamical zeta functions (22.11) factorize as

$$\frac{1}{\zeta} = \prod_{\mu} \frac{1}{\zeta_{\mu}^{d_{\mu}}}, \quad \frac{1}{\zeta_{\mu}} = \prod_{\hat{p}} \det \left(1 - D^{(\mu)}(g_{\hat{p}}) t_{\hat{p}} \right). \quad (25.20)$$

Determinants of d -dimensional irreps can be evaluated using the expansion of determinants in terms of traces,

$$\begin{aligned} \det(1 + M) &= 1 + \operatorname{tr} M + \frac{1}{2} \left((\operatorname{tr} M)^2 - \operatorname{tr} M^2 \right) \\ &\quad + \frac{1}{6} \left((\operatorname{tr} M)^3 - 3 (\operatorname{tr} M)(\operatorname{tr} M^2) + 2 \operatorname{tr} M^3 \right) \\ &\quad + \cdots + \frac{1}{d!} \left((\operatorname{tr} M)^d - \cdots \right), \end{aligned} \quad (25.21)$$

and each factor in (25.19) can be evaluated by looking up the characters $\chi^{(\mu)}(g) = \operatorname{tr} D^{(\mu)}(g)$ in standard tables [12]. In terms of characters, we have for the 1-dimensional representations

$$\det(1 - D^{(\mu)}(g)t) = 1 - \chi^{(\mu)}(g)t,$$

for the 2-dimensional representations

$$\det(1 - D^{(\mu)}(g)t) = 1 - \chi^{(\mu)}(g)t + \frac{1}{2} \left(\chi^{(\mu)}(g)^2 - \chi^{(\mu)}(g^2) \right) t^2,$$

and so forth.

In the fully symmetric subspace $\operatorname{tr} D_{A_1}(g) = 1$ for all orbits; hence a straightforward fundamental domain computation (with no group theory weights) always yields a part of the full spectrum. In practice this is the most interesting subspectrum, as it contains the leading eigenvalue of the evolution operator.

exercise 25.2

25.4.2 Factorization of spectral determinants

Factorization of the full spectral determinant (22.3) proceeds in essentially the same manner as the factorization of dynamical zeta functions outlined above. By (25.14) the trace of the evolution operator $\mathcal{L}^t$ splits into the sum of inequivalent irreducible subspace contributions $\sum_{\mu} \operatorname{tr} \mathcal{L}_{\mu}^t$, with

$$\operatorname{tr} \mathcal{L}_{\mu}^t = d_{\mu} \sum_{g \in G} \chi^{(\mu)}(g) \int_{\hat{\mathcal{M}}} d\hat{x} \mathcal{L}^t(D(g)^{-1} \hat{x}, \hat{x}).$$

This leads by standard manipulations to the factorization of (22.8) into

$$\begin{aligned} F(z) &= \prod_{\mu} F_{\mu}(z)^{d_{\mu}} \\ F_{\mu}(z) &= \exp \left(- \sum_{\hat{p}} \sum_{r=1}^{\infty} \frac{1}{r} \frac{\chi^{(\mu)}(g_{\hat{p}}^r) z^{n_{\hat{p}} r}}{|\det(\mathbf{1} - \hat{M}_{\hat{p}}^r)|} \right), \end{aligned} \quad (25.22)$$

where $\hat{M}_{\hat{p}} = D(g_{\hat{p}})M_{\hat{p}}$ is the fundamental domain Jacobian. Boundary orbits require special treatment, discussed in sect. 25.4.3, with examples given in the next section as well as in the specific factorizations discussed below.

25.4.3 Boundary orbits

(Continued from sect. 11.4)

Before we can turn to a presentation of the factorizations of dynamical zeta functions for the different symmetries we have to discuss an effect that arises for orbits that run on a symmetry line that borders a fundamental domain. In our 3-disk example, no such orbits are possible, but they exist in other systems, such as in the bounded region of the Hénon-Heiles potential and in 1-d maps. For the symmetrical 4-disk billiard, there are in principle two kinds of such orbits, one kind bouncing back and forth between two diagonally opposed disks and the other kind moving along the other axis of reflection symmetry; the latter exists for bounded systems only. While there are typically very few boundary orbits, they tend to be among the shortest orbits, and their neglect can seriously degrade the convergence of cycle expansions, as those are dominated by the shortest cycles.

While such orbits are invariant under some symmetry operations, their neighborhoods are not. This affects the Jacobian matrix M_p of the linearization perpendicular to the orbit and thus the eigenvalues. Typically, *e.g.* if the symmetry is a reflection, some eigenvalues of M_p change sign. This means that instead of a weight $1/\det(\mathbf{1} - M_p)$ as for a regular orbit, boundary cycles also pick up contributions of form $1/\det(\mathbf{1} - D(g)M_p)$, where $D(g)$ is a symmetry operation that leaves the orbit pointwise invariant; see example 25.9.

Consequences for the dynamical zeta function factorizations are that sometimes a boundary orbit does not contribute. A derivation of a dynamical zeta function (22.11) from a determinant like (22.8) usually starts with an expansion of the determinants of the Jacobian. The leading order terms just contain the product of the expanding eigenvalues and lead to the dynamical zeta function (22.11). Next to leading order terms contain products of expanding and contracting eigenvalues and are sensitive to their signs. Clearly, the weights t_p in the dynamical zeta function will then be affected by reflections in the Poincaré section perpendicular to the orbit. In all our applications it was possible to implement these effects by the following simple prescription.



If an orbit is invariant under a little group $G_p = \{e, b_2, \dots, b_h\}$, then the corresponding group element in (25.16) will be replaced by a projector. If the weights are insensitive to the signs of the eigenvalues, then this projector is

$$g_p = \frac{1}{h} \sum_{i=1}^h b_i. \quad (25.23)$$

In the cases that we have considered, the change of sign may be taken into account by defining a sign function $\epsilon_p(g) = \pm 1$, with the “-” sign if the symmetry element

g flips the neighborhood. Then (25.23) is replaced by

$$g_p = \frac{1}{h} \sum_{i=1}^h \epsilon(b_i) b_i. \quad (25.24)$$

The factorizations (25.20), (25.22) are the central formulas of this chapter. We now work out the group theory factorizations of cycle expansions of dynamical zeta functions for Z_2 and D_3 symmetries. D_2 and D_4 symmetries are worked out in appendix A25.

25.5 $Z_2 = D_1$ factorization

As the simplest example of implementing the above scheme consider the $Z_2 = D_1$ symmetry. For our purposes, all that we need to know here is that each orbit or configuration is uniquely labeled by an infinite string $\{s_i\}$, $s_i = +, -$ and that the dynamics is invariant under the $+ \leftrightarrow -$ interchange, i.e., it is Z_2 symmetric. The Z_2 symmetry cycles separate into two classes, the self-dual configurations $+-, ++--, +++---, +---+--+ , \dots$, with multiplicity $m_p = 1$, and the asymmetric configurations $+, -, ++-, --+, \dots$, with multiplicity (10.6) $m_p = 2$. For example, as there is no absolute distinction between the “up” and the “down” spins, or the “left” or the “right” lobe, $t_+ = t_-$, $t_{++} = t_{--}$, and so on.

exercise 25.6

The symmetry reduced labeling $\rho_i \in \{0, 1\}$ is related to the standard $s_i \in \{+, -\}$ Ising spin labeling by

$$\begin{aligned} \text{If } s_i &= s_{i-1} \text{ then } \rho_i = 1 \\ \text{If } s_i &\neq s_{i-1} \text{ then } \rho_i = 0 \end{aligned} \quad (25.25)$$

For example, $\overline{+} = \dots + + + \dots$ maps into $\dots 111 \dots = \overline{1}$ (and so does $\overline{-}$), $\overline{+-} = \dots - + - + \dots$ maps into $\dots 000 \dots = \overline{0}$, $\overline{++-} = \dots - - + + - - + + \dots$ maps into $\dots 0101 \dots = \overline{01}$, and so forth. A list of such reductions is given in table 15.1.

Depending on the maximal symmetry group G_p that leaves an orbit p invariant (see sect. 25.3 as well as example 25.9), the contributions to the dynamical zeta function factor as

$$\begin{aligned} G_p &= \{e\} : \quad (1 - t_{\hat{p}})^2 = (1 - t_{\hat{p}})(1 - t_{\hat{p}}) \\ G_p &= \{e, \sigma\} : \quad (1 - t_{\hat{p}}^2) = (1 - t_{\hat{p}})(1 + t_{\hat{p}}), \end{aligned} \quad (25.26)$$

For example:

$$\begin{aligned} G_{++-} &= \{e\} : \quad (1 - t_{++-})^2 = (1 - t_{001})(1 - t_{001}) \\ G_{+-} &= \{e, \sigma\} : \quad (1 - t_{+-}) = (1 - t_0)(1 + t_0), \quad t_{+-} = t_0^2 \end{aligned}$$

This yields two binary cycle expansions. The A_1 subspace dynamical zeta function is given by the standard binary expansion (23.8). The antisymmetric A_2 subspace dynamical zeta function ζ_{A_2} differs from ζ_{A_1} only by a minus sign for cycles with an odd number of 0's:

$$\begin{aligned}
 1/\zeta_{A_2} &= (1+t_0)(1-t_1)(1+t_{10})(1-t_{100})(1+t_{101})(1+t_{1000}) \\
 &\quad (1-t_{1001})(1+t_{1011})(1-t_{10000})(1+t_{10001}) \\
 &\quad (1+t_{10010})(1-t_{10011})(1-t_{10101})(1+t_{10111}) \dots \\
 &= 1+t_0-t_1+(t_{10}-t_1t_0)-(t_{100}-t_{10}t_0)+(t_{101}-t_{10}t_1) \\
 &\quad -(t_{1001}-t_1t_{001}-t_{101}t_0+t_{10}t_0t_1)-\dots \quad (25.27)
 \end{aligned}$$

Note that the group theory factors do not destroy the curvature corrections (the cycles and pseudo cycles are still arranged into shadowing combinations).

If the system under consideration has a boundary orbit (cf. sect. 25.4.3) with group-theoretic factor $h_p = (e + \sigma)/2$, the boundary orbit does not contribute to the antisymmetric subspace

$$\begin{array}{cc}
 A_1 & A_2 \\
 \text{boundary: } (1-t_p) &= (1-t_{\hat{p}})(1-0t_{\hat{p}})
 \end{array} \quad (25.28)$$

This is the $1/\zeta$ part of the boundary orbit factorization discussed in example 25.9, where the factorization of the corresponding spectral determinants for the 1-dimensional reflection symmetric maps is worked out in detail.



example 25.9
p. 520

25.6 D_3 factorization: 3-disk game of pinball

The next example, the D_3 symmetry, can be worked out by a glance at figure 15.12 (a). For the symmetric 3-disk game of pinball the fundamental domain is bounded by a disk segment and the two adjacent sections of the symmetry axes that act as mirrors (see figure 15.12 (b)). The three symmetry axes divide the space into six copies of the fundamental domain. Any trajectory on the full space can be pieced together from bounces in the fundamental domain, with symmetry axes replaced by flat mirror reflections. The binary $\{0, 1\}$ reduction of the ternary three disk $\{1, 2, 3\}$ labels has a simple geometric interpretation: a collision of type 0 reflects the projectile to the disk it comes from (back-scatter), whereas after a collision of type 1 projectile continues to the third disk. For example, $\overline{23} = \dots 232323 \dots$ maps into $\dots 000 \dots = \overline{0}$ (and so do $\overline{12}$ and $\overline{13}$), $\overline{123} = \dots 12312 \dots$ maps into $\dots 111 \dots = \overline{1}$ (and so does $\overline{132}$), and so forth. A list of such reductions for short cycles is given in table 15.2.

D_3 has two 1-dimensional irreps, symmetric and antisymmetric under reflections, denoted A_1 and A_2 , and a pair of degenerate 2-dimensional representations of mixed symmetry, denoted E . The contribution of an orbit with symmetry g to the $1/\zeta$ Euler product (25.19) factorizes according to

$$\det(1 - D^{reg}(h)t) = \left(1 - \chi^{(A_1)}(h)t\right) \left(1 - \chi^{(A_2)}(h)t\right) \left(1 - \chi^{(E)}(h)t + \chi^{(A_2)}(h)t^2\right)^2 \quad (25.29)$$

with the three factors contributing to the D_3 irreps A_1 , A_2 and E , respectively, and the 3-disk dynamical zeta function factorizes into $\zeta = \zeta_{A_1} \zeta_{A_2} \zeta_E^2$. Substituting the D_3 characters [12]

D_3	A_1	A_2	E
e	1	1	2
d, d_2	1	1	-1
σ_v	1	-1	0

into (25.29), we obtain for the three classes of possible orbit symmetries (indicated in the first column)

$$\begin{array}{lll} g_{\hat{p}} & A_1 & A_2 & E \\ e : & (1 - t_{\hat{p}})^6 & = & (1 - t_{\hat{p}})(1 - t_{\hat{p}})(1 - 2t_{\hat{p}} + t_{\hat{p}}^2)^2 \\ d, d_2 : & (1 - t_{\hat{p}}^3)^2 & = & (1 - t_{\hat{p}})(1 - t_{\hat{p}})(1 + t_{\hat{p}} + t_{\hat{p}}^2)^2 \\ \sigma_v : & (1 - t_{\hat{p}}^2)^3 & = & (1 - t_{\hat{p}})(1 + t_{\hat{p}})(1 + 0t_{\hat{p}} - t_{\hat{p}}^2)^2. \end{array} \quad (25.30)$$

where σ_v stands for any one of the three reflections.

The Euler product (22.11) on each irreducible subspace follows from the factorization (25.30). On the symmetric A_1 subspace the ζ_{A_1} is given by the standard binary curvature expansion (23.8). The antisymmetric A_2 subspace ζ_{A_2} differs from ζ_{A_1} only by a minus sign for cycles with an odd number of 0's, and is given in (25.27). For the mixed-symmetry subspace E the curvature expansion is given by

$$\begin{aligned} 1/\zeta_E &= (1 + zt_1 + z^2t_1^2)(1 - z^2t_0^2)(1 + z^3t_{100} + z^6t_{100}^2)(1 - z^4t_{10}^2) \\ &\quad (1 + z^4t_{1001} + z^8t_{1001}^2)(1 + z^5t_{10000} + z^{10}t_{10000}^2) \\ &\quad (1 + z^5t_{10101} + z^{10}t_{10101}^2)(1 - z^5t_{10011})^2 \dots \\ &= 1 + zt_1 + z^2(t_1^2 - t_0^2) + z^3(t_{001} - t_1t_0^2) \\ &\quad + z^4 \left[t_{0011} + (t_{001} - t_1t_0^2)t_1 - t_{01}^2 \right] \\ &\quad + z^5 \left[t_{00001} + t_{01011} - 2t_{00111} + (t_{0011} - t_{01}^2)t_1 + (t_1^2 - t_0^2)t_{100} \right] + \dots \end{aligned} \quad (25.31)$$

We have reinserted the powers of z in order to group together cycles and pseudo-cycles of the same length. Note that the factorized cycle expansions retain the

curvature form; long cycles are still shadowed by (somewhat less obvious) combinations of pseudo-cycles.

Referring back to the topological polynomial (18.41) obtained by setting $t_p = 1$, we see that its factorization is a consequence of the D_3 factorization of the ζ function:

$$1/\zeta_{A_1} = 1 - 2z, \quad 1/\zeta_{A_2} = 1, \quad 1/\zeta_E = 1 + z, \quad (25.32)$$

as obtained from (23.8), (25.27) and (25.31) for $t_p = 1$.

Their symmetry is $K = \{e, \sigma\}$, so according to (25.23), they pick up the group-theoretic factor $g_p = (e + \sigma)/2$. If there is no sign change in t_p , then evaluation of $\det(1 - \frac{e+\sigma}{2}t_{\hat{p}})$ yields

$$\text{boundary: } (1 - t_p)^3 = (1 - t_{\hat{p}})(1 - 0t_{\hat{p}})(1 - t_{\hat{p}})^2, \quad t_p = t_{\hat{p}}. \quad (25.33)$$

However, if the cycle weight changes sign under reflection, $t_{\sigma\hat{p}} = -t_{\hat{p}}$, the boundary orbit does not contribute to the subspace symmetric under reflection across the orbit;

$$\text{boundary: } (1 - t_p)^3 = (1 - 0t_{\hat{p}})(1 - t_{\hat{p}})(1 - t_{\hat{p}})^2, \quad t_p = t_{\hat{p}}. \quad (25.34)$$

Résumé

If a dynamical system has a discrete symmetry, the symmetry should be exploited; much is gained, both in understanding of the spectra and ease of their evaluation. Once this is appreciated, it is hard to conceive of a calculation without factorization; it would correspond to quantum mechanical calculations without wave-function symmetrizations.



While the reformulation of the chaotic spectroscopy from the trace sums to the cycle expansions does not reduce the exponential growth in number of cycles with the cycle period, in practice only the short orbits are used, and for them the labor saving is dramatic. For example, for the 3-disk game of pinball there are 256 periodic points of period 8, but reduction to the fundamental domain non-degenerate prime cycles reduces the number of the distinct cycles of period 8 to 30.

In addition, cycle expansions of the symmetry reduced dynamical zeta functions converge dramatically faster than the unfactorized dynamical zeta functions. One reason is that the unfactorized dynamical zeta function has many closely spaced zeros and zeros of multiplicity higher than one; since the cycle expansion

is a polynomial expansion in topological cycle period, accommodating such behavior requires many terms. The dynamical zeta functions on separate subspaces have more evenly and widely spaced zeros, are smoother, do not have symmetry-induced multiple zeros, and fewer cycle expansion terms (short cycle truncations) suffice to determine them. Furthermore, the cycles in the fundamental domain sample state space more densely than in the full space. For example, for the 3-disk problem, there are 9 distinct (symmetry unrelated) cycles of period 7 or less in full space, corresponding to 47 distinct periodic points. In the fundamental domain, we have 8 (distinct) periodic orbits up to period 4 and thus 22 different periodic points in $1/6$ -th the state space, i.e., an increase in density by a factor 3 with the same numerical effort.

We emphasize that the symmetry factorization (25.30) of the dynamical zeta function is *intrinsic* to the deterministic dynamics, and not a special property of quantal spectra. The factorization is not restricted to the Hamiltonian systems, or only to the configuration space symmetries; for example, the discrete symmetry can be a symmetry of the Hamiltonian phase space [16]. In conclusion, the manifold advantages of the symmetry reduced dynamics should thus be obvious; full state space cycle expansions, such as those of exercise 25.4, are useful only for cross-checking purposes.



Commentary

Remark 25.1. Symmetry reductions in periodic orbit theory. Some of the standard references on characters and irreps of compact groups are refs. [2, 6, 8, 12, 21]. We found Tinkham [19] introduction to the basic concepts the most enjoyable.

This chapter is based on a collaborative effort with B. Eckhardt. The group-theoretic factorizations of dynamical zeta functions that we develop here were first introduced and applied in ref. [3]. They are closely related to the symmetrizations introduced by Gutzwiller [10] in the context of the semiclassical periodic orbit trace formulas, put into more general group-theoretic context by Robbins [16], whose exposition, together with Lauritzen's [13] treatment of the boundary orbits, has influenced the presentation given here. The symmetry reduced trace formula for a finite symmetry group $G = \{e, g_2, \dots, g_{|G|}\}$ with $|G|$ group elements, where the integral over Haar measure is replaced by a finite group discrete sum $|G|^{-1} \sum_{g \in G} = 1$, derived in ref. [3]. A related group-theoretic decomposition in context of hyperbolic billiards was utilized in ref. [1], and for the Selberg's zeta function in ref. [20]. One of its loftier antecedents is the Artin factorization formula of algebraic number theory, which expresses the zeta-function of a finite extension of a given field as a product of L -functions over all irreps of the corresponding Galois group.

The techniques of this chapter have been applied to computations of the 3-disk deterministic and quantum spectra in refs. [7, 17], and to a "Zeeman effect" pinball and the x^2y^2 potentials in ref. [4, 5]. In a larger perspective, the factorizations developed above are special cases of a general approach to exploiting the group-theoretic invariances in spectra computations, such as those used in enumeration of periodic geodesics [15, 18] for hyperbolic billiards [9] and Selberg zeta functions [11].

References

- [1] N. L. Balász and A. Voros, “Chaos on the pseudosphere”, *Phys. Rep.* **143**, 109–240 (1986).
- [2] J. F. Cornwell, *Group Theory in Physics: An Introduction* (Academic, New York, 1997).
- [3] P. Cvitanović and B. Eckhardt, “Symmetry decomposition of chaotic dynamics”, *Nonlinearity* **6**, 277–311 (1993).
- [4] P. Dahlqvist, “Determination of resonance spectra for bound chaotic systems”, *J. Phys. A* **27**, 763–785 (1994).
- [5] P. Dahlqvist and G. Russberg, “Periodic orbit quantization of bound chaotic systems”, *J. Phys. A* **24**, 4763–4778 (1991).
- [6] M. S. Dresselhaus, G. Dresselhaus, and A. Jorio, *Group Theory: Application to the Physics of Condensed Matter* (Springer, New York, 2007).
- [7] B. Eckhardt, G. Russberg, P. Cvitanović, P. E. Rosenqvist, and P. Scherer, “Pinball scattering”, in *Quantum Chaos: Between Order and Disorder*, edited by G. Casati and B. Chirikov (Cambridge Univ. Press, Cambridge UK, 1995).
- [8] J. P. Elliott and P. G. Dawber, *Symmetry in Physics*, Vol. 2 (MacMillan, Surrey, 1979).
- [9] M. C. Gutzwiller, “Phase-integral approximation in momentum space and the bound states of an atom”, *J. Math. Phys.* **8**, 1979–2000 (1967).
- [10] M. C. Gutzwiller, “The quantization of a classically ergodic system”, *Physica D* **5**, 183–207 (1982).
- [11] M. C. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer, New York, 1990).
- [12] M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Dover, New York, 1962).
- [13] B. Lauritzen, “Discrete symmetries and the periodic-orbit expansions”, *Phys. Rev. A* **43**, 603–606 (1991).
- [14] E. N. Lorenz, “Deterministic nonperiodic flow”, *J. Atmos. Sci.* **20**, 130–141 (1963).
- [15] H. P. McKean, “Selberg’s trace formula as applied to a compact Riemann surface”, *Commun. Pure Appl. Math.* **25**, 225–246 (1972).
- [16] J. M. Robbins, “Discrete symmetries in periodic-orbit theory”, *Phys. Rev. A* **40**, 2128–2136 (1989).
- [17] P. Scherer, Quantenzustände eines Klassisch Chaotischen Billards, PhD thesis (Univ. Köln, 1991).
- [18] M. Sieber and F. Steiner, “Classical and quantum mechanics of a strongly chaotic billiard system”, *Physica D* **44**, 248–266 (1990).
- [19] M. Tinkham, *Group Theory and Quantum Mechanics* (Dover, New York, 2003).

- [20] A. B. Venkov and P. G. Zograf, “On analogues of the Artin factorization formulas in the spectral theory of automorphic functions connected with induced representations of Fuchsian groups”, *Math. USSR* **21**, 435–443 (1983).
- [21] E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1931).

25.7 Examples

Example 25.1. A matrix representation of 2-element group $\mathbf{Z}_2$: If a 2-dimensional map $f(x)$ has the symmetry $x_1 \rightarrow -x_1, x_2 \rightarrow -x_2$, the symmetry group G consists of the identity and d , a rotation by π around the origin. The map f must then commute with rotations by π , $f(D(d)x) = D(d)f(x)$, with the matrix representation of d given by the $[2 \times 2]$ matrix

$$D(d) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (25.35)$$

d satisfies $d^2 = e$ and can be used to decompose the state space into mutually orthogonal symmetric and antisymmetric subspaces by means of projection operators (25.50).
(continued in example 25.3)

[return to p. 500](#)

Example 25.2. A matrix representation of cyclic group $\mathbf{C}_3$: A 3-dimensional matrix representation of the 3-element cyclic group $\mathbf{C}_3 = \{e, d, d_2\}$ is given by the three rotations by $2\pi/3$ around z -axis in a 3-dimensional state space,



$$\begin{aligned} D(e) &= \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}, \quad D(d) = \begin{bmatrix} \cos \frac{2\pi}{3} & -\sin \frac{2\pi}{3} & \\ \sin \frac{2\pi}{3} & \cos \frac{2\pi}{3} & \\ & & 1 \end{bmatrix}, \\ D(d_2) &= \begin{bmatrix} \cos \frac{4\pi}{3} & -\sin \frac{4\pi}{3} & \\ \sin \frac{4\pi}{3} & \cos \frac{4\pi}{3} & \\ & & 1 \end{bmatrix}. \end{aligned}$$

(continued in example 25.4)

(X. Ding)

[return to p. 500](#)

Example 25.3. A 2-tiles state space: The state space $\mathcal{M} = \{x_1-x_2 \text{ plane}\}$ of example 25.1, with symmetry group $G = \{e, d\}$, can be tiled by a fundamental domain $\hat{\mathcal{M}} = \{\text{half-plane } x_1 \geq 0\}$, and $d\hat{\mathcal{M}} = \{\text{half-plane } x_1 \leq 0\}$, its image under rotation by π .

[return to p. 501](#)

Example 25.4. The regular representation of cyclic group $\mathbf{C}_3$: (continued from example 25.2) Take an arbitrary function $\rho(x)$ over the state space $x \in \mathcal{M}$, and define a fundamental domain $\hat{\mathcal{M}}$ as a $1/3$ wedge, with axis z as its (symmetry invariant) edge. The state space is tiled with three copies of the wedge,

$$\mathcal{M} = \hat{\mathcal{M}}_1 \cup \hat{\mathcal{M}}_2 \cup \hat{\mathcal{M}}_3 = \hat{\mathcal{M}} \cup d\hat{\mathcal{M}} \cup d_2\hat{\mathcal{M}}.$$

Function $\rho(x)$ can be written as the 3-dimensional vector of functions over the fundamental domain $\hat{x} \in \hat{\mathcal{M}}$,

$$(\rho_1^{reg}(\hat{x}), \rho_2^{reg}(\hat{x}), \rho_3^{reg}(\hat{x})) = (\rho(\hat{x}), \rho(d\hat{x}), \rho(d_2\hat{x})). \quad (25.36)$$

The multiplication table of $\mathbf{C}_3$ is given in table 25.2. By (25.5), the regular representation matrices $D^{reg}(g)$ have '1' at the location of g in the multiplication table, '0' elsewhere. The actions of the operator $U(g)$ are now represented by permutations matrices (blank entries are zeros):

$$D^{reg}(e) = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}, \quad D^{reg}(d) = \begin{bmatrix} & 1 & \\ & & 1 \\ 1 & & \end{bmatrix}, \quad D^{reg}(d_2) = \begin{bmatrix} & & 1 \\ 1 & & \\ & 1 & \end{bmatrix}. \quad (25.37)$$

(X. Ding)

[return to p. 502](#)

Z_2	e	σ
e	e	σ
σ^{-1}	σ	e

C_3	e	d	d_2
e	e	d	d_2
d_{-1}	d_2	e	d
d_{-2}	d	d_2	e

Table 25.2: The multiplication tables of the 2-element group Z_2 , and C_3 , the group of symmetries of a 3-blade propeller.

Z_2	e	σ
A	1	1
B	1	-1

C_3	e	d	d_2
0	1	1	1
1	1	ω	ω^2
2	1	ω^2	ω

D_3	e	3σ	$2d$
A	1	1	1
B	1	-1	1
E	2	0	-1

Table 25.3: Z_2 , C_3 and D_3 character tables. The classes $\{\sigma_{12}, \sigma_{13}, \sigma_{14}\}$, $\{d, d_2\}$ are denoted 3σ , $2d$, respectively.

Example 25.5. The regular representation of dihedral group D_3 : The multiplication table of D_3 is given in table 25.1. By (25.5), the 6 regular representation matrices $D^{reg}(g)$ have ‘1’ at the location of g in the multiplication table, ‘0’ elsewhere. For example, the regular representation of the action of operators $U(\sigma_{23})$ and $U(d_2)$ are given in table 25.1.

(X. Ding)

[return to p. 502](#)

Example 25.6. Irreps of cyclic group C_3 : (continued from example 25.4) We would like to generalize the symmetric-antisymmetric functions decomposition of C_2 to the order 3 group C_3 . Symmetrization can be carried out on any number of functions, but there is no obvious ‘antisymmetrization’. We draw instead inspiration from the Fourier transformation for a finite periodic lattice, and construct from the regular basis (25.36) a new set of basis functions

$$\rho_0(\hat{x}) = \frac{1}{3} [\rho(\hat{x}) + \rho(d\hat{x}) + \rho(d_2\hat{x})] \quad (25.38)$$

$$\rho_1(\hat{x}) = \frac{1}{3} [\rho(\hat{x}) + \omega \rho(d\hat{x}) + \omega^2 \rho(d_2\hat{x})] \quad (25.39)$$

$$\rho_2(\hat{x}) = \frac{1}{3} [\rho(\hat{x}) + \omega^2 \rho(d\hat{x}) + \omega \rho(d_2\hat{x})]. \quad (25.40)$$

The representation of group C_3 in this new basis is block diagonal by inspection:

$$D(e) = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}, \quad D(d) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega^2 \end{bmatrix}, \quad D(d_2) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \omega^2 & 0 \\ 0 & 0 & \omega \end{bmatrix}. \quad (25.41)$$

Here $\omega = e^{2i\pi/3}$. So C_3 has three 1-dimensional irreps ρ_0 , ρ_1 and ρ_2 . Generalization to any C_n is immediate: this is just a finite lattice Fourier transform.

(X. Ding)

[return to p. 504](#)

Example 25.7. Character table of D_3 : (continued from example 25.5) Let us construct table 25.3. Spectroscopists conventions are to use labels A and B for symmetric, respectively antisymmetric nondegenerate irreps, and E for the doubly degenerate irreps. So 1-dimensional representations are denoted by A and B , depending on whether the basis function is symmetric or antisymmetric with respect to transpositions σ_{ij} . E denotes the

2-dimensional representation. As D_3 has 3 classes, the dimension sum rule $d_1^2 + d_2^2 + d_3^2 = 6$ has only one solution $d_1 = d_2 = 1, d_3 = 2$. Hence there are two 1-dimensional irreps and one 2-dimensional irrep. The first column is 1, 1, 2, and the first row is 1, 1, 1 corresponding to the 1-d symmetric representation. We take two approaches to figure out the remaining 4 entries. First, since B is antisymmetric 1-d representation, so the characters should be ± 1 . We anticipate $\chi^{(B)}(\sigma) = -1$ and can quickly figure out the remaining 3 positions. We check the obtained table satisfies the orthonormal relations. Second, denote $\chi^{(B)}(\sigma) = x$ and $\chi^{(E)}(\sigma) = y$, then from the orthonormal relation of the second column with the first column and itself, we obtain $1 + x + 2y = 0$, and $1 + x^2 + y^2 = 6/3$, we get two sets of solutions, one of them can be shown not compatible with other orthonormality relations, so $x = -1, y = 0$. Similarly, we can get the other two characters. (X. Ding)

[return to p. 504](#)

Example 25.8. Basis for irreps of D_3 : (continued from example 25.7) From table 25.3, we have



$$P^A \rho(x) = \frac{1}{6} [\rho(x) + \rho(\sigma_{12}x) + \rho(\sigma_{23}x) + \rho(\sigma_{31}x) + \rho(dx) + \rho(d_2x)] \quad (25.42)$$

$$P^B \rho(x) = \frac{1}{6} [\rho(x) - \rho(\sigma_{12}x) - \rho(\sigma_{23}x) - \rho(\sigma_{31}x) + \rho(dx) + \rho(d_2x)] \quad (25.43)$$

For projection into irrep E, we need to figure out the explicit matrix representation first. Obviously, the following 2 by 2 matrices are E irrep.

$$D^E(e) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad D^E(d) = \begin{bmatrix} \omega & 0 \\ 0 & \omega^2 \end{bmatrix}, \quad D^E(d_2) = \begin{bmatrix} \omega^2 & 0 \\ 0 & \omega \end{bmatrix} \quad (25.44)$$

$$D^E(\sigma_{12}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad D^E(\sigma_{23}) = \begin{bmatrix} 0 & \omega^2 \\ \omega & 0 \end{bmatrix}, \quad D^E(\sigma_{31}) = \begin{bmatrix} 0 & \omega \\ \omega^2 & 0 \end{bmatrix} \quad (25.45)$$

So apply projection operator on $\rho(x)$ and $\rho(\sigma_{12}x)$:

$$P_1^E \rho(x) = \frac{1}{3} [\rho(x) + \omega \rho(dx) + \omega^2 \rho(d_2x)] \quad (25.46)$$

$$P_2^E \rho(x) = \frac{1}{3} [\rho(x) + \omega^2 \rho(dx) + \omega \rho(d_2x)] \quad (25.47)$$

$$P_1^E \rho(\sigma_{12}x) = \frac{1}{3} [\rho(\sigma_{12}x) + \omega \rho(\sigma_{31}x) + \omega^2 \rho(\sigma_{23}x)] \quad (25.48)$$

$$P_2^E \rho(\sigma_{12}x) = \frac{1}{3} [\rho(\sigma_{12}x) + \omega^2 \rho(\sigma_{31}x) + \omega \rho(\sigma_{23}x)] \quad (25.49)$$

Under the invariant basis

$$\{P^A \rho(x), P^B \rho(x), P_1^E \rho(x), P_2^E \rho(\sigma_{12}x), P_1^E \rho(\sigma_{12}x), P_2^E \rho(x)\}$$

,

$$D(\sigma_{23}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega^2 & 0 & 0 \\ 0 & 0 & \omega & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \omega^2 \\ 0 & 0 & 0 & 0 & \omega & 0 \end{bmatrix} \quad D(d) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \omega & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \omega & 0 \\ 0 & 0 & 0 & 0 & 0 & \omega^2 \end{bmatrix}.$$

(X. Ding)

[return to p. 504](#)

Example 25.9. Reflection symmetric 1- d maps: Consider f , a map on the interval with reflection symmetry $f(-x) = -f(x)$. A simple example is the piecewise-linear sawtooth map of figure 11.4. Denote the reflection operation by $\sigma x = -x$. The symmetry of the map implies that if $\{x_n\}$ is a trajectory, then also $\{\sigma x_n\}$ is a trajectory because $\sigma x_{n+1} = \sigma f(x_n) = f(\sigma x_n)$. The dynamics can be restricted to a fundamental domain, in this case to one half of the original interval; every time a trajectory leaves this interval, it can be mapped back using σ . Furthermore, the evolution operator is invariant under the group, $U(\sigma)\mathcal{L}^t(y, x) = \mathcal{L}^t(y, x)$. σ satisfies $\sigma^2 = \mathbf{e}$ and can be used to decompose the state space into mutually orthogonal symmetric and antisymmetric subspaces by means of projection operators

$$\begin{aligned} P_{A_1} &= \frac{1}{2}(1 + U(\sigma)), & P_{A_2} &= \frac{1}{2}(1 - U(\sigma)), \\ \mathcal{L}_{A_1}^t(y, x) &= P_{A_1}\mathcal{L}^t(y, x) = \frac{1}{2}(\mathcal{L}^t(y, x) + \mathcal{L}^t(-y, x)), \\ \mathcal{L}_{A_2}^t(y, x) &= P_{A_2}\mathcal{L}^t(y, x) = \frac{1}{2}(\mathcal{L}^t(y, x) - \mathcal{L}^t(-y, x)). \end{aligned} \quad (25.50)$$

To compute the traces of the symmetrization and antisymmetrization projection operators (25.50), we have to distinguish three kinds of cycles: asymmetric cycles a , symmetric cycles s built by repeats of irreducible segments $\tilde{s}$, and boundary cycles b . Now we show that the spectral determinant can be written as the product over the three kinds of cycles: $\det(1 - \mathcal{L}') = \det(1 - \mathcal{L}')_a \det(1 - \mathcal{L}')_{\tilde{s}} \det(1 - \mathcal{L}')_b$.

Asymmetric cycles: A periodic orbits is not symmetric if $\{x_a\} \cap \{\sigma x_a\} = \emptyset$, where $\{x_a\}$ is the set of periodic points belonging to the cycle a . Thus σ generates a second orbit with the same number of points and the same stability properties. Both orbits give the same contribution to the first term and no contribution to the second term in (25.50); as they are degenerate, the prefactor 1/2 cancels. Resuming as in the derivation of (22.11) we find that asymmetric orbits yield the same contribution to the symmetric and the antisymmetric subspaces:

$$\det(1 - \mathcal{L}_{\pm})_a = \prod_a \prod_{k=0}^{\infty} \left(1 - \frac{t_a}{\Lambda_a^k}\right), \quad t_a = \frac{z_a^{n_a}}{|\Lambda_a|}.$$

Symmetric cycles: A cycle s is reflection symmetric if operating with σ on the set of periodic points reproduces the set. The period of a symmetric cycle is always even ($n_s = 2n_{\tilde{s}}$) and the mirror image of the x_s periodic point is reached by traversing the irreducible segment $\tilde{s}$ of length $n_{\tilde{s}}$, $f^{n_{\tilde{s}}}(x_s) = \sigma x_s$. $\delta(x - f^n(x))$ picks up $2n_{\tilde{s}}$ contributions for every even traversal, $n = rn_{\tilde{s}}$, r even, and $\delta(x + f^n(x))$ for every odd traversal, $n = rn_{\tilde{s}}$, r odd. Absorb the group-theoretic prefactor in the Floquet multiplier by defining the stability computed for a segment of length $n_{\tilde{s}}$,

$$\Lambda_{\tilde{s}} = - \left. \frac{\partial f^{n_{\tilde{s}}}(x)}{\partial x} \right|_{x=x_s}.$$

Restricting the integration to the infinitesimal neighborhood $\mathcal{M}_s$ of the s cycle, we obtain

the contribution to $\text{tr } \mathcal{L}_\pm^n$:

$$\begin{aligned} z^n \text{tr } \mathcal{L}_\pm^n &\rightarrow \int_{\mathcal{M}_s} dx z^n \frac{1}{2} (\delta(x - f^n(x)) \pm \delta(x + f^n(x))) \\ &= n_{\bar{s}} \left(\sum_{r=2}^{\text{even}} \delta_{n, rn_{\bar{s}}} \frac{t_{\bar{s}}^r}{1 - 1/\Lambda_{\bar{s}}^r} \pm \sum_{r=1}^{\text{odd}} \delta_{n, rn_{\bar{s}}} \frac{t_{\bar{s}}^r}{1 - 1/\Lambda_{\bar{s}}^r} \right) \\ &= n_{\bar{s}} \sum_{r=1}^{\infty} \delta_{n, rn_{\bar{s}}} \frac{(\pm t_{\bar{s}})^r}{1 - 1/\Lambda_{\bar{s}}^r}. \end{aligned}$$

Substituting all symmetric cycles s into $\det(1 - \mathcal{L}_\pm)$ and resumming we obtain:

$$\det(1 - \mathcal{L}_\pm)_{\bar{s}} = \prod_{\bar{s}} \prod_{k=0}^{\infty} \left(1 \mp \frac{t_{\bar{s}}}{\Lambda_{\bar{s}}^k} \right)$$

Boundary cycles: In the example at hand there is only one cycle which is neither symmetric nor antisymmetric, but lies on the boundary of the fundamental domain, the fixed point at the origin. Such cycle contributes simultaneously to both $\delta(x - f^n(x))$ and $\delta(x + f^n(x))$:

$$\begin{aligned} z^n \text{tr } \mathcal{L}_\pm^n &\rightarrow \int_{\mathcal{M}_b} dx z^n \frac{1}{2} (\delta(x - f^n(x)) \pm \delta(x + f^n(x))) \\ &= \sum_{r=1}^{\infty} \delta_{n,r} t_b^r \frac{1}{2} \left(\frac{1}{1 - 1/\Lambda_b^r} \pm \frac{1}{1 + 1/\Lambda_b^r} \right) \\ z^n \text{tr } \mathcal{L}_+^n &\rightarrow \sum_{r=1}^{\infty} \delta_{n,r} \frac{t_b^r}{1 - 1/\Lambda_b^{2r}}; \quad z^n \text{tr } \mathcal{L}_-^n \rightarrow \sum_{r=1}^{\infty} \delta_{n,r} \frac{1}{\Lambda_b^r} \frac{t_b^r}{1 - 1/\Lambda_b^{2r}}. \end{aligned}$$

Boundary orbit contributions to the factorized spectral determinants follow by resumming:

$$\det(1 - \mathcal{L}_+) = \prod_{k=0}^{\infty} \left(1 - \frac{t_b}{\Lambda_b^{2k}} \right), \quad \det(1 - \mathcal{L}_-) = \prod_{k=0}^{\infty} \left(1 - \frac{t_b}{\Lambda_b^{2k+1}} \right)$$

Only the even derivatives contribute to the symmetric subspace, and only the odd ones to the antisymmetric subspace, because the orbit lies on the boundary.

Finally, the symmetry reduced spectral determinants follow by collecting the above results:

$$\begin{aligned} F_+(z) &= \prod_a \prod_{k=0}^{\infty} \left(1 - \frac{t_a}{\Lambda_a^k} \right) \prod_{\bar{s}} \prod_{k=0}^{\infty} \left(1 - \frac{t_{\bar{s}}}{\Lambda_{\bar{s}}^k} \right) \prod_{k=0}^{\infty} \left(1 - \frac{t_b}{\Lambda_b^{2k}} \right) \\ F_-(z) &= \prod_a \prod_{k=0}^{\infty} \left(1 - \frac{t_a}{\Lambda_a^k} \right) \prod_{\bar{s}} \prod_{k=0}^{\infty} \left(1 + \frac{t_{\bar{s}}}{\Lambda_{\bar{s}}^k} \right) \prod_{k=0}^{\infty} \left(1 - \frac{t_b}{\Lambda_b^{2k+1}} \right) \end{aligned} \quad (25.51)$$

exercise 25.1
return to p. 511

Example 25.10. 3-disk billiard / D_3 cycle weights factorized: Compare, for example, the contributions of the $\overline{12}$ and $\overline{0}$ cycles of figure 15.12. $\text{tr } D^{reg}(h) \hat{\mathcal{L}}$ does not

get a contribution from the $\bar{0}$ cycle, as the symmetry operation that maps the first half of the $\bar{12}$ into the fundamental domain is a reflection, and $\text{tr } D^{\text{reg}}(\sigma) = 0$. In contrast, $\sigma^2 = e$, $\text{tr } D^{\text{reg}}(\sigma^2) = 6$ insures that the repeat of the fundamental domain fixed point $\text{tr } (D^{\text{reg}}(h)\hat{\mathcal{L}})^2 = 6t_0^2$, gives the correct contribution to the global trace $\text{tr } \mathcal{L}^2 = 3 \cdot 2t_{12}$.

We see by inspection in figure 15.12 that $t_{12} = t_0^2$ and $t_{123} = t_1^3$.

[return to p. 506](#)

25.8 Exercises

- 25.1. **Sawtooth map desymmetrization.** Work out the some of the shortest global cycles of different symmetries and fundamental domain cycles for the sawtooth map of figure 11.4. Compute the dynamical zeta function and the spectral determinant of the Perron-Frobenius operator for this map; check explicitly the factorization (25.51).
- 25.2. **2-dimensional asymmetric representation.** The above expressions can sometimes be simplified further using standard group-theoretical methods. For example, the $\frac{1}{2}((\text{tr } M)^2 - \text{tr } M^2)$ term in (25.21) is the trace of the antisymmetric part of the $M \times M$ Kronecker product. Show that if α is a 2-dimensional representation, this is the A_2 antisymmetric representation, and

$$2\text{-dim: } \det(1 - D_\alpha(h)t) = 1 - \chi_\alpha(h)t + \chi_{A_2}(h)t^2. \quad (25.52)$$

- 25.3. **Characters of D_3 .** (continued from exercise 10.5) D_3 , the group of symmetries of an equilateral triangle: has three irreducible representations, two one-dimensional and the other one of multiplicity 2.
- (a) All finite discrete groups are isomorphic to a permutation group or one of its subgroups, and elements of the permutation group can be expressed as cycles. Express the elements of the group D_3 as cycles. For example, one of the rotations is (123), meaning that vertex 1 maps to 2, 2 $\rightarrow$ 3, and 3 $\rightarrow$ 1.
 - (b) Use your representation from exercise 10.5 to compute the D_3 character table.
 - (c) Use a more elegant method from the group-theory literature to verify your D_3 character table.
 - (d) Two D_3 irreducible representations are one dimensional and the third one of multiplicity 2 is formed by $[2 \times 2]$ matrices. Find the matrices for all six group elements in this representation.

(Hint: get yourself a good textbook, like Hamermesh [12] or Tinkham [19], and read up on classes and characters.)

- 25.4. **3-disk unfactorized zeta cycle expansions.** Check that the curvature expansion (23.3) for the 3-disk pinball, assuming no symmetries between disks, is given by

$$\begin{aligned} 1/\zeta &= (1 - z^2 t_{12})(1 - z^2 t_{13})(1 - z^2 t_{23}) \\ &\quad (1 - z^3 t_{123})(1 - z^3 t_{132})(1 - z^4 t_{1213}) \\ &\quad (1 - z^4 t_{1232})(1 - z^4 t_{1323})(1 - z^5 t_{12123}) \cdots \\ &= 1 - z^2 t_{12} - z^2 t_{23} - z^2 t_{31} - z^3 (t_{123} + t_{132}) \\ &\quad - z^4 [(t_{1213} - t_{12} t_{13}) + (t_{1232} - t_{12} t_{23}) \\ &\quad + (t_{1323} - t_{13} t_{23})] \\ &\quad - z^5 [(t_{12123} - t_{12} t_{123}) + \cdots] - \cdots \end{aligned} \quad (25.53)$$

Show that the symmetrically arranged 3-disk pinball cycle expansion of the Euler product

(23.3) (see table 18.5 and figure 10.1) is given by:

$$\begin{aligned}
 1/\zeta &= (1 - z^2 t_{12})^3 (1 - z^3 t_{123})^2 (1 - z^4 t_{1213})^3 \\
 &\quad (1 - z^5 t_{12123})^6 (1 - z^6 t_{121213})^6 \\
 &\quad (1 - z^6 t_{121323})^3 \dots \\
 &= 1 - 3z^2 t_{12} - 2z^3 t_{123} - 3z^4 (t_{1213} - t_{12}^2) \\
 &\quad - 6z^5 (t_{12123} - t_{12} t_{123}) \\
 &\quad - z^6 (6 t_{121213} + 3 t_{121323} + t_{12}^3 - 9 t_{12} t_{1213} - t_{123}^2) \\
 &\quad - 6z^7 (t_{1212123} + t_{1212313} + t_{1213123} + t_{12}^2 t_{123} \\
 &\quad - 3 t_{12} t_{12123} - t_{123} t_{1213}) \\
 &\quad - 3z^8 (2 t_{12121213} + t_{12121313} + 2 t_{12121323} \\
 &\quad + 2 t_{12123123} + 2 t_{12123213} + t_{12132123} \\
 &\quad + 3 t_{12}^2 t_{1213} + t_{12} t_{123}^2 - 6 t_{12} t_{121213} \\
 &\quad - 3 t_{12} t_{121323} - 4 t_{123} t_{12123} - t_{123}^2) - \dots
 \end{aligned} \tag{25.54}$$

25.5. 3-disk desymmetrization.

- Work out the 3-disk symmetry factorization for the 0 and 1 cycles, i.e. which symmetry do they have, what is the degeneracy in full space and how do they factorize (how do they look in the A_1 , A_2 and the E representations).
- Find the shortest cycle with no symmetries and factorize it as in a)
- Find the shortest cycle that has the property that its time reversal is not described by the same symbolic dynamics.
- Compute the dynamical zeta functions and the spectral determinants (symbolically) in the three representations; check the factorizations (25.20) and (25.22).

(Per Rosenqvist)

25.6. C_2 factorizations: the Lorenz and Ising systems. In the Lorenz system [14] the labels + and - stand for the left or the right lobe of the attractor and the symmetry is a rotation by π around the z -axis. Similarly, the Ising Hamiltonian (in the absence of an external magnetic field) is invariant under spin flip. Work out the factorizations for some of the short cycles in either system.

25.7. Ising model. The Ising model with two states $\epsilon_i = \{+, -\}$ per site, periodic boundary condition, and Hamiltonian

$$H(\epsilon) = -J \sum_i \delta_{\epsilon_i, \epsilon_{i+1}},$$

is invariant under spin-flip: $+\leftrightarrow -$. Take advantage of that symmetry and factorize the dynamical zeta function for the model, i.e., find all the periodic orbits that contribute to each factor and their weights.

25.8. One orbit contribution. If p is an orbit in the fundamental domain with symmetry h , show that it contributes to the spectral determinant with a factor

$$\det \left(1 - D^{reg}(h) \frac{t_p}{\lambda_p^k} \right),$$

where $D^{reg}(h)$ is the regular representation of G .

Chapter 26

Continuous symmetry factorization

Hard work builds character.

— V.I. [Warszawski](#), Private Investigator

TRACE FORMULAS relate short time dynamics (unstable periodic orbits) to long time invariant state space densities (natural measure). Higher dimensional dynamics requires inclusion of higher-dimensional compact invariant sets, such as partially hyperbolic invariant tori, into trace formulas. A trace formula for a partially hyperbolic $(N + 1)$ -dimensional compact manifold invariant under N global continuous symmetries is derived here. In this extension of ‘periodic orbit’ theory there are no or very few periodic orbits - the relative periodic orbits that the trace formula has support on are almost never eventually periodic.

The deterministic trace formula for smooth continuous time flows

chapter 21

$$\sum_{\alpha=0}^{\infty} \frac{1}{s - s_{\alpha}} = \sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r(\beta A_p - s T_p)}}{|\det(\mathbf{1} - M_p^r)|}$$

relates the spectrum of the evolution operator

$$\mathcal{L}(x', x) = \delta(x' - f^t(x)) e^{\beta A(x,t)} \quad (26.1)$$

to the unstable periodic orbits p of the flow $f^t(x)$. This formula (and the associated spectral determinants and cycle expansions) is valid for fully hyperbolic flows.

chapter 22

Here we derive the corresponding formula for dynamics invariant under a compact group of symmetry transformations. In what follows, a familiarity with basic group-theoretic notions is assumed, with the definitions relegated to appendix [A10.1](#).



fast track:
chapter 30, p. 609

26.1 Compact groups

All the group theory that we shall need here is given by

The Peter-Weyl Theorem, and its corollaries: A compact Lie group G is completely reducible, its representations are fully reducible (just as in the finite group representation theory), every compact Lie group is a closed subgroup of $U(n)$ for some n , and every continuous, unitary, irreducible representation of a compact Lie group is finite dimensional.

The theory of semisimple Lie groups is elegant, perhaps too elegant. In what follows, we serve group theoretic nuggets in need-to-know portions, offering a pedestrian route through a series of simple examples of familiar aspects of group theory and Fourier analysis, and a high, cyclist road in the text proper.

But main idea is this: the character $\chi^{(m)}(\theta)$ of the Frobenius-Weyl representation theory is a generalization to all compact continuous Lie groups of the weight $e^{i\theta m}$ in the Fourier decomposition of a smooth function on a circle into eigenmodes of translation. m th Fourier component fits m node function around the circle; $(m_1, m_2, \dots, m_N)$ representation of a compact Lie group fits a corresponding multi-mode function onto the smooth manifold swept out by the action of the group. So a basis for a d -dimensional representation $(m_1, m_2, \dots, m_N)$ of an N -dimensional compact Lie group is a set of d linearly independent eigenfunctions on the N -dimensional compact group manifold, with $m_1, m_2, \dots, m_N$ ‘nodes’ along the N directions needed to span the manifold. For a circle this is Fourier analysis; for a sphere these are spherical harmonics, and the Peter-Weyl theorem states that analogous expansion exists for every compact Lie group. We will never need to construct these explicitly.

exercise 26.2

26.1.1 Group representations

Let q_a be a vector in d -dimensional vector space V , and G be a group of linear transformations

$$q'_a = D(g)_a{}^b q_b, \quad a, b = 1, 2, \dots, d, \quad g \in G$$

(repeated indices summed throughout this chapter). The $[d \times d]$ matrices $D(g)$ form a representation of the group G . Vectors in the dual space $\bar{q}$ transform as

$$q'^a = D(g)^a{}_b q^b.$$

Tensors transform as

$$h'^{ab} = D(g)^a{}_f D(g)^b{}_e D(g)^c{}_d h_f e^d.$$

A function H is an invariant function if (and only if) for any transformation $g \in G$ and for any set of vectors $\bar{q}, \bar{r}, s, \dots$

$$H(D(g)^\dagger \bar{q}, D(g)^\dagger \bar{r}, \dots, D(g)s) = H(\bar{q}, \bar{r}, \dots, s). \quad (26.2)$$

Unitary transformations connected to the identity can be generated by sequences of infinitesimal transformations

$$D(g)_a^b \simeq \delta_a^b + i\epsilon_i (T_i)_a^b \quad \epsilon_i \in \mathbb{R}, \quad T_i \text{ hermitian},$$

and $|\epsilon_i| \ll 1$. The N group generators T_a , $a = 1, \dots, N$ close the Lie algebra of G . (More generally, one also needs to study invariance under discrete coordinate transformations (see chapter 25).

Consider a multilinear invariant function

$$H(\bar{q}, \bar{r}, \dots, s) = h_{ab\dots}^{c\dots} q^a r^b \dots s_c$$

In terms of the generators T_i , H is invariant if all generators “annihilate” it, $T_i \cdot h = 0$:

$$(T_i)_a^{a'} h_{a'b\dots}^{c\dots} + (T_i)_b^{b'} h_{ab'\dots}^{c\dots} - (T_i)_c^{c'} h_{ab\dots}^{c'\dots} + \dots = 0. \quad (26.3)$$

Vector space V is irreducible if the only invariant subspaces of V under the action of G are (0) and V . If every V on which G acts can be written as a direct sum of irreducible subspaces, then G is completely reducible.



example 26.1
p. 538

26.1.2 Group integrals

Consider a group integral of form

$$\int dg D(g)_a^b D(g)_d^c, \quad (26.4)$$

where $D(g)_a^b$ is a unitary $[d \times d]$ matrix representation of $g \in G$, G a compact Lie group, $D(g)_d^c$ is the matrix representation of the action of g on the dual vector space,

$$D(g)_d^c = (D(g)^\dagger)_d^c,$$

and the integration is over the entire range of $g \in G$, G a compact Lie group. For a finite group G with $|G|$ group elements the normalized measure is a discrete sum,

$$d\mu(x) = \frac{1}{|G|} \sum_g \delta(gx).$$

For continuous groups, the integration measure dg is known as the Haar measure, and, given an explicit parametrization of the group manifold, is explicitly computable (see example 26.4 and example 26.5). However, we do need such explicit parametrizations, as the integral (26.4) over the entire group is defined by two requirements:

exercise 26.1

1. *Normalization:* The group average of an scalar quantity is the quantity itself,

$$\int dg = 1. \quad (26.5)$$

2. *Orthonormality of irreducible representations.* How do we define

$$\int dg D(g)_a^b = ?$$

The action of $g \in G$ is to rotate a vector x_a into $x'_a = D(g)_a^b x_b$

$$\begin{array}{ccc} & g & \\ x' & & x \end{array}$$

Manifold traced out by action of G
for all possible group elements g

The averaging smears x in all directions, hence the second integration rule

$$\int dg D(g)_a^b = 0, \quad \text{if } D(g) \text{ is non-trivial representation,} \quad (26.6)$$

simply states that the average over all rotations of a vector is zero.

A representation is trivial (a ‘singlet’) if $D(g) = 1$ for all group elements g . In this case no averaging is taking place, and the first integration rule (26.5) applies.

What happens if we average a bilinear combination of a pair of vectors x, y ? There is no reason why such pair should average to zero; for example, we know that the scalar function $|x|^2 = \sum_a x_a x_a^* = x_a x^a$ is invariant under unitary transformations, so it cannot have a vanishing average. Therefore, in general

$$\int dg D(g)_a^b D(g)^c_d \neq 0. \quad (26.7)$$

To get a feeling for what the right-hand side looks like, we recommend that you work out the examples.

 example 26.2
p. 538

Now let $D(g)$ be any irreducible $[d \times d]$ rep. Irreducibility (known in this context as ‘Schur’s Lemma’) means that any invariant $[d \times d]$ tensor A_b^a is proportional to δ_b^a . As the only bilinear invariant is δ_b^a , the Clebsch-Gordan series

$$\begin{array}{c} \leftarrow \\ \rightarrow \end{array} = \frac{1}{d} \begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} + \sum_{\lambda}^{\text{irreps}} \begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \begin{array}{c} \lambda \\ \curvearrowright \end{array} \begin{array}{c} \curvearrowleft \end{array} \quad (26.8)$$

contains one and only one singlet. Only the singlet survives the group averaging, and

$$\int dg D^{(\lambda)}(g)_a^d D^{(\lambda)}(g)_c^b = \frac{1}{d} \delta_c^d \delta_a^b. \quad (26.9)$$

is true for any $[d \times d]$ irreducible rep $D^{(\lambda)}(g)$.

If we take $D^{(\mu)}(g)_a^\beta$ and $D^{(\lambda)}(g)_d^c$ in inequivalent representations λ, μ (there is no matrix K such that $D^{(\lambda)}(g) = K D^{(\mu)}(g) K^{-1}$ for any $g \in G$), then there is no way of forming a singlet, and

$$\int dg D^{(\lambda)}(g)_a^d D^{(\mu)}(g)_c^\beta = 0 \quad \text{if} \quad \lambda \neq \mu. \quad (26.10)$$

26.1.3 Characters

The trace of an irreducible $[d \times d]$ matrix representation λ of g is called the *character* of the representation:

$$\chi^{(\lambda)}(g) = \text{tr } D^{(\lambda)}(g) = D^{(\lambda)}(g)_a^a. \quad (26.11)$$

The character of the conjugate representation is

$$\chi^{(\lambda)}(g^{-1}) = \text{tr } D^{(\lambda)}(g)^\dagger = D^{(\lambda)}(g)_a^a = \chi^{(\lambda)}(g)^*. \quad (26.12)$$

Contracting (26.8) with two arbitrary invariant $[d \times d]$ tensors h_a^a and $(f^\dagger)_b^c$, we obtain the *character orthonormality relation*

$$\int dg \chi^{(\lambda)}(hg) \chi^{(\mu)}(gf) = \delta_{\lambda\mu} \frac{1}{d_\lambda} \chi^{(\lambda)}(hf^\dagger) \quad (26.13)$$

The character orthonormality tells us that if two group invariant quantities share a $D^{(\lambda)}(g)D^{(\lambda)}(g^{-1})$ pair, the group averaging sews them into a single group invariant quantity. The replacement of $D^{(\lambda)}(g)_a^b$ by the character $\chi^{(\lambda)}(h^{-1}g)$ does not mean that the matrix structure is lost; $D^{(\lambda)}(g)_a^b$ can be recovered by differentiating

$$D(g)_a^b = \frac{d}{dh_b^a} \chi^{(\lambda)}(h^{-1}g). \quad (26.14)$$

The essential group theory we shall need here is most compactly summarized by

The Group Orthogonality Theorem: Let $D_\mu, D_{\mu'}$ be two irreducible matrix representations of a compact group G of dimensions $d_\mu, d_{\mu'}$,

$$\int dg D^{(\mu)}(g)_a^b D^{(\mu')}(g^{-1})^{a'}_{b'} = \frac{1}{d_\mu} \delta_{\mu,\mu'} \delta_a^{a'} \delta_{b'}^b.$$

The new trace formula follows from the full reducibility of representations of a compact group G acting linearly on a vector space V , with irreducible representations labeled by sets of integers $\mu = (\mu_1, \dots, \mu_N)$, and the vector space V

decomposed into invariant subspaces V_μ . For a N -dimensional compact Lie group G the fundamental result is the Weyl full reducibility theorem, with projection operator onto the V_μ irreducible subspace given by

$$P_\mu = d_\mu \int_G g \chi^{(\mu)}(g^{-1}) U(g). \quad (26.15)$$

The group elements $g = g(\theta_1, \dots, \theta_N) = e^{i\theta \cdot T}$ are parameterized by N real numbers $\{\theta_1, \dots, \theta_N\}$ of finite range, hence designation ‘compact’.

26.1.4 Transformation operators, projection operators

Suppose we have an arbitrary function or set of functions. How do we obtain functions with desired symmetry properties? If f is an arbitrary function,

$$P_{ij}^\alpha f(x) = \frac{d_\alpha}{|G|} \sum_G D^{(\alpha)}(-1) g f(x) = F_{ij}^\alpha(x)$$

which is either zero or a basis function for the i th row of irrep α : a function of symmetry species (α, i) .



example 26.3
p. 539



example 26.4
p. 539



example 26.5
p. 539

The character χ is the trace $\chi^{(\mu)}(g) = \text{tr } D_\mu(g) = \sum_{i=1}^{d_\mu} D_\mu(g)_{ii}$, where $D_\mu(g)$ is a $[d_\mu \times d_\mu]$ -dimensional matrix representation of action of the group element g on the irreducible subspace V_μ . We will sometimes employ notation g as a shorthand for $D(g)$, i.e., by $x' = gx$ we mean the matrix operation $x'_i = \sum_{j=1}^d D(g)_{ij} x_j$, and by $f'(x) = g f(x) = f(gx)$, $f(x)$ a smooth function over the state space $x \in \mathcal{M}$, we mean $f'(x) = f(D(g)x)$.

For an invariant scalar quantity the average over the group in (26.15) must be the quantity itself, so the group integral is weighted by the normalized Haar measure (unit group volume) $\int dg = 1$, and d_μ is the multiplicity of degenerate eigenvalues in representation μ .

26.2 Continuous symmetries of dynamics

If action of every element g of a compact group G commutes with the flow $\dot{x} = v(x)$,

$$D(g)v(x) = v(D(g)x), \quad D(g)f^t(x) = f^t(D(g)x),$$

G is a global symmetry of the dynamics. The finite time evolution operator (26.1) can be written as $\mathcal{L}^t = e^{t\mathcal{A}}$ in terms of the time-evolution generator (19.25)

$$\mathcal{A} = \lim_{\delta\tau \rightarrow 0^+} \frac{1}{\delta\tau} (\mathcal{L}^{\delta\tau} - I) = -\partial_i(v_i(x)). \quad (26.16)$$

The operator $e^{t\mathcal{A}}$ commutes with all symmetry transformations $e^{i\theta \cdot T}$. For a given state space point x together they sweep out a $(N+1)$ -dimensional manifold of equivalent orbits.

As in (25.13), $\mathcal{L}(y, x)$ is invariant function. The irreducible eigenspaces of G are also eigenspaces of the dynamical evolution operator $\mathcal{L}^t$, with the decomposition of the evolution operator to irreducible subspaces, $\mathcal{L} = \sum_{\mu} \mathcal{L}_{\mu}$, following immediately by application of the projection operator (26.15):

$$\mathcal{L}_{\mu}^t(y, x) = d_{\mu} \int_G dg \chi^{(\mu)}(g) \mathcal{L}^t(D_{\mu}(g^{-1})y, x). \quad (26.17)$$

As G commutes with f^t , all eigenfunctions ρ of $\mathcal{L}^t$ must be invariant under G , $\rho(x) = \rho(gx)$. Infinitesimally, in terms of Lie algebra generators $T_{\phi}\rho(x) = 0$.

26.2.1 Relative periodic orbits

Relative periodic orbits are orbits $x(t)$ in state space $\mathcal{M}$ which exactly recur

$$x(t) = D(g_p^r) x(t + rT_p) \quad (26.18)$$

for a fixed *relative period* T , its repeats $r = 1, 2, \dots$, and a fixed group action $g \in G$ of $\mathcal{M}$. This group action is sometimes referred to as a ‘phase’, or a ‘shift’. Relative periodic orbits are to periodic solutions what relative equilibria (traveling waves) are to equilibria (steady solutions).

For dynamical systems with continuous symmetries relative periodic orbits are almost never eventually periodic, i.e., they almost never lie on periodic trajectories in the full state space. As almost any such orbit explores ergodically the manifold swept by action of G , they are sometimes referred to as ‘quasiperiodic.’ However, an orbit can be periodic if it satisfies a special symmetry. If $g^m = 1$ is of finite order m , then the corresponding orbit is periodic with period mT . If g is not of finite order k , orbits can be periodic only after the action of g .

chapter 25

In either case, we refer to the orbits in $\mathcal{M}$ satisfying (26.18) as *relative periodic orbits*.

26.2.2 Stability of relative periodic orbits

A infinitesimal group transformation maps globally a trajectory in a nearby trajectory, so we expect the initial point perturbations along to group manifold to be marginal, growing at rates slower than exponential. The argument is akin to (4.9), the proof of marginality of perturbations along the trajectory. Consider two nearby initial points separated by an infinitesimal group rotation $\delta\theta$: $\delta x_0 = f^{\delta\theta}(x_0) - x_0 = v(x_0)\delta\theta$. By the commutativity of the group with the flow, $f^{t+\delta t} = f^{\delta t} \circ f^t$. Expanding both sides of $f^t(f^{\delta t}(x_0)) = f^{\delta t}(f^t(x_0))$, keeping the leading term in δt , and using the definition of the Jacobian matrix (4.5), we observe

that $J^t(x_0)$ transports the velocity vector at x_0 to the velocity vector at $x(t)$ at time t :

$$v(x(t)) = J^t(x_0) v(x_0). \quad (26.19)$$

In nomenclature of page 88, the Jacobian matrix maps the initial, Lagrangian coordinate frame into the current, Eulerian coordinate frame.

However, already at this stage we see that if the orbit is periodic, $g_p x(T_p) = x(0)$, at any point along cycle p the velocity v is an eigenvector of the Jacobian matrix $J_p = J^{T_p}$ with an eigenvalue of unit magnitude,

$$J_p(x) v(x) = v(x), \quad x \in \mathcal{M}_p. \quad (26.20)$$

Two successive points along the cycle separated by δx_0 have the same separation after a completed period $\delta x(T_p) = g_p \delta x_0$, hence eigenvalue of magnitude 1.

26.3 Symmetry reduced trace formula for flows

As any pair of nearby points on a periodic orbit returns to itself exactly at each cycle period, the eigenvalue of the Jacobian matrix corresponding to the eigenvector along the flow necessarily equals unity for all periodic orbits. In presence of N -dimensional symmetry Lie group G , further N eigenvalues equal unity. Hence the trace integral $\text{tr } \mathcal{L}^t$ requires a separate treatment for the direction along the flow and for the N group transformation directions.

To evaluate the contribution of a prime cycle p of period T_p , restrict the integration to an infinitesimally thin manifold $\mathcal{M}_p$ enveloping the cycle and all of its rotations by G , pick a point on the cycle, and choose a local coordinate system with a longitudinal coordinate $dx_{\parallel}$ along the direction of the flow, N coordinates dx_G along the invariant manifold swept by p under the action of the symmetry group G , and $(d-N-1)$ transverse coordinates $x_{\perp}$.

$$\text{tr } {}_p \mathcal{L}_{\mu}^t = d_{\mu} \int_G g \chi^{(\mu)}(g) \int_{\mathcal{M}_p} dx_{\perp} dx_{\parallel} dx_G \delta(x - D(g)) f^t(x). \quad (26.21)$$

The integral along the longitudinal, time-evolution coordinate was computed in (21.16). Eliminating the time dependence by Laplace transform one obtains

$$\int_0^{\infty} e^{-st} \oint_p dx_{\parallel} \delta_{\parallel}(x_{\parallel} - f^t(x_{\parallel})) = T_p \sum_{r=1}^{\infty} e^{-sT_p r}. \quad (26.22)$$

 example 26.6
p. 539

The μ subspace group integral is simple:

$$\int_G g \chi^{(\mu)}(g) \int_{\mathcal{M}_p} dx_G \delta(x_G - D_{\mu}(g) f^{rT_p}(x_G)) = \chi^{(\mu)}(g_p^r). \quad (26.23)$$

For the remaining transverse coordinates the Jacobian matrix is defined in a $(N+1)$ -dimensional surface of section $\mathcal{P}$ of constant $(x_{\parallel}, x_G)$. Linearization of the periodic flow transverse to the orbit yields

$$\int_{\mathcal{P}} dx_{\perp} \delta(x_{\perp} - D_{\mu}(g_p^r) f^{rT_p}(x_{\perp})) = \frac{1}{|\det(\mathbf{1} - \hat{M}_p^r)|}, \quad (26.24)$$

where $\hat{M}_p = D_{\mu}(g_p)M_p$ is the p -cycle $[(d-1-N) \times (d-1-N)]$ symmetry reduced Jacobian matrix, computed on the reduced surface of section and rotated by g_p . As in (21.5), we assume hyperbolicity, i.e., that the magnitudes of all transverse eigenvalues are bounded away from unity.

The *deterministic symmetry reduced trace formula for flows* follows by substituting (26.22) - (26.24) into (26.21):

$$\sum_{\beta=0}^{\infty} \frac{1}{s - s_{\mu\beta}} = d_{\mu} \sum_p T_p \sum_{r=1}^{\infty} \chi^{(\mu)}(g_p^r) \frac{e^{r(\beta A_p - s T_p)}}{|\det(\mathbf{1} - \hat{M}_p^r)|}. \quad (26.25)$$

(we can restore $e^{\beta A_p}$ from (26.1) provided that the observable $a(x)$ also commutes with G .) The sum is over all prime *relative periodic orbits* p and their repeats, orbits in state space which satisfy

$$x(t) = D(g_p)x(t + T_p) \quad (26.26)$$

for a fixed *relative period* T_p and a fixed *shift* g_p .

The $\mu = (0, 0, \dots, 0)$ subspace is the one of most relevance to chaotic dynamics, as its leading eigenfunction, with the fewest nodes and the slowest decay rate, corresponds to the natural measure observed in the long time dynamics.

In contrast to the case of continuous symmetries, where relative periodic orbits are almost never eventually periodic, i.e., they almost never lie on periodic trajectories in the full state space, for discrete symmetries all relative periodic orbits are eventually periodic.

Résumé

One of the goals of nonlinear dynamics is to describe the long time evolution of ensembles of trajectories, when individual trajectories are exponentially unstable. The main tool in this effort have been trace formulas because they make explicit the duality between individual short time trajectories, and long time invariant densities (natural measures, eigenfunctions of evolution operators). So far, the main successes have been in applications to low dimensional flows and iterated mappings, where the compact invariant sets of short-time dynamics are equilibria, periodic points and periodic orbits. Dynamics in higher dimensions requires extension of trace formulas to higher-dimensional compact invariant sets, such as partially hyperbolic invariant tori.

Here we have used a particularly simple direct product structure of a global symmetry that commutes with the flow to reduce the dynamics to a symmetry reduced $(d-1-N)$ -dimensional state space M/G . The trace formulas do not require explicit construction (in general difficult), neither of the reduced state space, nor of the Haar measures.

Amusingly, in this extension of ‘periodic orbit’ theory from unstable 1-dimensional closed orbits to unstable $(N+1)$ -dimensional compact manifolds invariant under continuous symmetries, there are no or very few periodic orbits. Relative periodic orbits are almost never eventually periodic, i.e., they almost never lie on periodic trajectories in the full state space, unless forced to do so by a discrete symmetry, so looking for periodic orbits in systems with continuous spatial symmetries is a fool’s errand.

Restriction to compact Lie groups in derivation of the trace formula (26.25) was a matter of convenience, as the general case is more transparent than particular implementations (such as $SO(2)$ and $SO(3)$ rotations, with their explicit Haar measures and characters). This can be relaxed as the need arises - much powerful group theory developed since Cartan-Weyl era is at our disposal. For example, the time evolution is in general non-compact (a generic trajectory is an orbit of infinite length). Nevertheless, the trace formulas have support on compact invariant sets in M , such as periodic orbits and $(N+1)$ -dimensional manifolds generated from them by action of the global symmetry groups. Just as existence of a periodic orbit is a consequence of given dynamics, not any global symmetry, higher-dimensional flows beckon us on with nontrivial higher-dimensional compact invariant sets (for example, partially hyperbolic invariant tori) for whom the trace formulas are still to be written.

Commentary

Remark 26.1. Literature Here we need only basic results, on the level of any standard group theory textbook [11]. This material is covered in any introduction to linear algebra [9, 16, 19] We found Tinkham [25] the most enjoyable as a no-nonsense, the user friendliest introduction to the basic concepts. The construction of projection operators given here is taken from refs. [5–7]. Who wrote this down first we do not know, but we like Harter’s exposition [12–14] best. Harter’s theory of class algebras offers a more elegant and systematic way of constructing the maximal set of commuting invariant matrices M_i than the sketch offered here. Bluman and Kumei [1] Chapter 2 offers a clear and pedagogical introduction to Lie groups of transformations. For the Group Orthogonality Theorem see, for example, refs. [7, 27], or consult Googlette.



Remark 26.2. Full reducibility of semisimple Lie groups: The study of integrals over compact Lie groups with respect to Haar measure is important in many areas of mathematics and physics, see Mehta [17]. In 1896-1897 Frobenius introduced notions of ‘characters’ and group ‘representations’, and proved the full reducibility of representations of finite groups. The characters $\chi^{(\mu)}(g)$ for all compact semisimple Lie groups were constructed and the full reducibility proven by Weyl [21], extending Cartan’s local Lie

algebra classification to a global theory of group representations. For the history of this period, see the excellent essay by Hawkins [15].

Diagrammatic notation for group theory is explained in the birdtracks.eu webbook [7].

Remark 26.3. William G. Harter is a prodigy who at age 16 taught himself group theory by reading Hamermesh [11]. Bill was a graduate student at Caltech (1964-65), together with Ron Fox. They hated the atmosphere there and the teaching was terrible (Feynman did not teach that year but Harter and Feynman were good friends). Harter and Fox shared an interest in group theory and discovered that most of the group theory books in the physics library had been checked out in 1960-62 by Gell-Mann, Zweig and Glashow. That only half of the entering students were meant to complete their PhD's there led to lots of ugly competition. Harter transferred to UC Irvine, and, upon graduation, got a job at USC in LA. After a few years he suggested in a faculty meeting that the way they could improve their quality as a department was "to get rid of all the old farts." These same "old farts" soon voted to deny him tenure. He ended up in Campinas, Brazil. Fox rescued him from there by bringing him for an interview at Georgia Tech, where he was hired in late 1970's. He was brilliant, an asset for teaching, making all sorts of demonstration devices. He built a giant rotating table upon which he placed billiard balls, a wonderful demonstration of mechanical analogues for charged particle motion in crossed E and B fields. Everyone (except for one nefarious character) liked him, his work, and especially his devices. The faculty unanimously supported his promotion to tenure. He did not, however, think much of the Director of School of Physics, and made that clear. After an argument with the Director, he stormed out, mortally offended. Denied tenure again, in 1985 he moved to University of Arkansas, where he developed his [Soft Elegant Educational tools](#). You can follow his lectures [on line](#).

In 1987 Harter and Weeks [26] used Harter's theory of the rotational dynamics of molecules to calculate the rotational-vibrational spectra of the soccer ball-shaped molecule Buckminsterfullerene, C₆₀, or "buckyball." C₆₀ had been proposed in 1985 by chemists, who had seen a mass-spectra peak of atomic mass 720. By 1989 the Harter theory calculations led to a realization that chemists had been making C₆₀ since the early 1970s. In 1992 Science named C₆₀ "Molecule of the Year," and in 1996 Curl, Kroto and Smalley were awarded the Nobel Prize in Chemistry for their discovery of fullerenes.

Remark 26.4. A brief history of relativity: In context of semiclassical quantization Creagh and Littlejohn [3, 4] concentrate on the case when the continuous symmetry family of orbits includes a true periodic orbit (they use infinitesimal variation around true periodic orbit), not the symmetry reduced case considered here (where almost every relative periodic orbit of the symmetry-reduced dynamics is not a periodic orbit in the full space). They emphasize generalized surface of section dynamics. They refer to relative periodic orbits as 'generalized periodic orbits', with 'generalized period' $T_p = (T_p, \mu_p)$. They mention, but do not go to irreducible reps of the symmetry groups, hence no 'deterministic symmetry reduced trace formula for flows' (26.25) in these papers. Instead, they explicitly compute group volumes. In addition to the reduced dynamics weight $|\det(1 - M_\perp)|$ they get $\partial\theta/\partial J$ which we do not have. The Berkeley group did it right for discrete symmetries [22, 23].

Here we follow Creagh [2], and in the axially-symmetric case ref. [20]. Creagh refers to relative periodic orbits as 'pseudoperiodic' orbits. Ref. [20] refers to relative periodic orbits as 'reduced periodic' orbits, and to the corresponding orbits in the full state space as 'quasiperiodic'. Creagh remarks at the very end of his paper to his formula (6.4) as the "pleasing result that the quantally reduced spectrum is determined by the deterministically reduced periodic orbits in the usual way." Ref. [10] discusses a trace formula in symmetry-reduced space. Muratore-Ginanneschi [18] gives an elegant discussion of

‘zero-modes’ in the path integral formulation, but does not go to irreps either. Ref. [24] applies the method to the problems of noninteracting identical particles.

References

- [1] G. W. Bluman and S. Kumei, *Symmetries and Differential Equations* (Springer, New York, 1989).
- [2] S. C. Creagh, “Semiclassical mechanics of symmetry reduction”, *J. Phys. A* **26**, 95–118 (1993).
- [3] S. C. Creagh and R. G. Littlejohn, “Semiclassical trace formulas in the presence of continuous symmetries”, *Phys. Rev. A* **44**, 836–850 (1991).
- [4] S. C. Creagh and R. G. Littlejohn, “Semiclassical trace formulas for systems with non-Abelian symmetry”, *J. Phys. A* **25**, 1643–1669 (1992).
- [5] P. Cvitanović, “Group theory for Feynman diagrams in non-Abelian gauge theories”, *Phys. Rev. D* **14**, 1536–1553 (1976).
- [6] P. Cvitanović, *Classical and exceptional Lie algebras as invariance algebras*, Oxford Univ. preprint 40/77, unpublished., 1977.
- [7] P. Cvitanović, *Group Theory: Birdtracks, Lie’s and Exceptional Groups* (Princeton Univ. Press, Princeton NJ, 2008).
- [8] J. P. Elliott and P. G. Dawber, *Symmetry in Physics*, Vol. 2 (MacMillan, Surrey, 1979).
- [9] I. M. Gel’fand, *Lectures on Linear Algebra* (Dover, New York, 1961).
- [10] V. Guillemin and A. Uribe, “Circular symmetry and the trace formula”, *Inv. Math.* **96**, 385–423 (1989).
- [11] M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Dover, New York, 1962).
- [12] W. G. Harter, “Algebraic theory of ray representations of finite groups”, *J. Math. Phys.* **10**, 739–752 (1969).
- [13] W. G. Harter, *Principles of Symmetry, Dynamics, and Spectroscopy* (Wiley, New York, 1993).
- [14] W. G. Harter and N. dos Santos, “Double-group theory on the half-shell and the two-level system. I. Rotation and half-integral spin states”, *Amer. J. Phys.* **46**, 251–263 (1978).
- [15] T. Hawkins, *Emergence of the Theory of Lie Groups: An Assay in the History of Mathematics* (Springer, New York, 2000), pp. 1869–1926.
- [16] S. Lang, *Linear Algebra* (Addison-Wesley, Reading, MA, 1987).
- [17] M. L. Mehta, *Random Matrices* (Academic, New York, 2004).
- [18] P. Muratore-Ginanneschi, “Path integration over closed loops and Gutzwiller’s trace formula”, *Phys. Rep.* **383**, 299–397 (2002).
- [19] K. Nomizu, *Fundamentals of Linear Algebra* (Chelsea Publ., New York, 1979).

- [20] S. Pal and D. Biswas, “Symmetry reduction and semiclassical analysis of axially symmetric systems”, *Phys. Rev. E* **57**, 1475–1484 (1998).
- [21] F. Peter and H. Weyl, “Die Vollständigkeit der primitiven Darstellungen einer geschlossenen kontinuierlichen Gruppe”, *Math. Ann.* **97**, 737–755 (1927).
- [22] J. M. Robbins, “Discrete symmetries in periodic-orbit theory”, *Phys. Rev. A* **40**, 2128–2136 (1989).
- [23] J. M. Robbins, S. C. Creagh, and R. G. Littlejohn, “Complex periodic orbits in the rotational spectrum of molecules: The example of SF₆”, *Phys. Rev. A* **39**, 2838–2854 (1989).
- [24] J. Sakhr and N. D. Whelan, “Semiclassical trace formulas for noninteracting identical particles”, *Phys. Rev. E* **67**, 066213 (2003).
- [25] M. Tinkham, *Group Theory and Quantum Mechanics* (Dover, New York, 2003).
- [26] D. E. Weeks and W. G. Harter, “Vibrational frequencies and normal modes of buckminsterfullerene”, *Chem. Phys. Lett.* **144**, 366–372 (1988).
- [27] E. W. Weisstein, *Group orthogonality theorem*, 2012.

26.4 Examples

Example 26.1. Lie algebra. As one does not want the rules to change at every step, the generators T_i are themselves invariant tensors,

$$(T_i)_b^a = D(g)_{a'}^a D(g)_b^{b'} D^{(A)}(g)_{i' i} (T_i)_{b'}^{a'}, \quad (26.27)$$

where $D^{(A)}(g)_{ij}$ is the adjoint $[N \times N]$ matrix representation of $g \in G$. For infinitesimal transformations, $D(g)_a^b \simeq \delta_a^b + i\epsilon_i (T_i)_a^b$. The $[d \times d]$ matrices T_i are in general non-commuting, and from (26.3) it follows that they close N -element *Lie algebra*

$$T_i T_j - T_j T_i = i C_{ijk} T_k \quad i, j, k = 1, 2, \dots, N,$$

where C_{ijk} are the *structure constants*.

[return to p. 527](#)

Example 26.2. A group integral for $SU(n) \bar{V} \times V$ space. Let $D(g)$ be the defining $[n \times n]$ matrix representation of $SU(n)$. The defining representation is non-trivial, so it averages to zero by (26.6). The first non-vanishing average involves $D(g)^\dagger$, the matrix representation of the action of g on the conjugate vector space. To avoid dealing with the multitude of dummy indices, we resort to diagrammatic notation:

$$D(g)_a^\ell = a \leftarrow \triangleleft \leftarrow \ell, \quad D(g)_\ell^a = a \rightarrow \triangleleft \rightarrow \ell. \quad (26.28)$$

For G the arrows and the triangle point the same way, while for $G^\dagger$ they point the opposite way. Unitarity $D(g)^\dagger D(g) = 1$ is given by

$$D(g)_a^c D(g)_c^b = D(g)_a^c D(g)_c^b = \delta_a^b,$$

or, diagrammatically:

$$\leftarrow \triangleright \leftarrow \triangleleft \leftarrow = \leftarrow \triangleleft \leftarrow \triangleright \leftarrow = \leftarrow \leftarrow. \quad (26.29)$$

In this notation, the $D(g)D(g)^\dagger$ integral (26.7) to be evaluated is

$$\int dg \begin{array}{c} a \leftarrow \triangleleft \leftarrow \epsilon \\ b \rightarrow \triangleleft \rightarrow \epsilon \end{array}. \quad (26.30)$$

For $SU(n)$ the $V \otimes \bar{V}$ tensors decompose into the singlet and the adjoint rep

$$\begin{array}{c} \leftarrow \\ \rightarrow \end{array} = \frac{1}{n} \begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} + \frac{1}{a} (T_i)_a^b (T_i)_c^d.$$

We multiply (26.30) with the above decomposition of the identity. The unitarity relation (26.29) eliminates G 's from the singlet:

$$\begin{array}{c} \leftarrow \triangleleft \leftarrow \\ \rightarrow \triangleleft \rightarrow \end{array} = \frac{1}{n} \begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} + \begin{array}{c} \leftarrow \triangleleft \leftarrow \\ \rightarrow \triangleleft \rightarrow \end{array}. \quad (26.31)$$

The generators T_i are invariant tensors, and transform under G according to (26.27). Multiplying by G_{ii}^{-1} , we obtain

$$\begin{array}{c} \leftarrow \triangleleft \leftarrow \\ \rightarrow \triangleleft \rightarrow \end{array} = \begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array}. \quad (26.32)$$

Hence, the pair $GG^\dagger$ in the defining representation can be traded in for a single G in the adjoint rep

$$\begin{aligned} D(g)_a^d D(g)^b_c &= \frac{1}{d} \delta_c^d \delta_a^b + \frac{1}{a} (T_i)_a^b G_{ij} (T_j)_c^d \\ \begin{array}{c} \text{---} \triangleleft \text{---} \\ \text{---} \triangleright \text{---} \end{array} &= \frac{1}{n} \begin{array}{c} \text{---} \curvearrowright \text{---} \\ \text{---} \curvearrowleft \text{---} \end{array} + \begin{array}{c} \text{---} \triangleleft \text{---} \\ \text{---} \triangleright \text{---} \end{array} \end{aligned}$$

The adjoint representation G_{ij} is non-trivial, so it gets averaged to zero by (26.6). Only the singlet survives

$$\begin{aligned} \int dg \begin{array}{c} \text{---} \triangleleft \text{---} \\ \text{---} \triangleright \text{---} \end{array} &= \frac{1}{d} \begin{array}{c} \text{---} \curvearrowright \text{---} \\ \text{---} \curvearrowleft \text{---} \end{array} \\ \int dg D(g)_a^d D(g)^b_c &= \frac{1}{d} \delta_c^d \delta_a^b. \end{aligned} \quad (26.33)$$

[return to p. 528](#)

Example 26.3. Irreducible representations of the $SO(2)^N$ abelian group: (Example 12.5 continued) All irreducible representations of the $SO(2)^N$ abelian group acting on torus T^N are 1-dimensional and labeled by N integers $\mu = (m_1, \dots, m_N)$. The character of μ representation is

$$\chi^{(\mu)}(g) = e^{-i\mu \cdot \phi}$$

[return to p. 530](#)

Example 26.4. Haar measure for $SO(2)$:

The normalized Haar measure is $dg = d\phi/(2\pi)$.

[return to p. 530](#)

Example 26.5. Haar measure for $SO(3)$:

$$SO(3) : \quad dg = \frac{1}{2\pi^2} \sin^2(\phi/2) d\Omega_e d\phi$$

with $d\Omega_e$ solid angle surface element for unit vector e .

$$8\pi^2 = \int_{SO(3)} dg$$

For details, see ref. [8].

[return to p. 530](#)

Example 26.6. Trace group integral for $SO(2)$: Parameterize rotations on a circle by $\phi \in [0, 2\pi)$. The normalized Haar measure is $ddg = d\phi/2\pi$, and a trajectory point advanced by time t and shifted by ϕ can be denoted $x(t, \phi)$. The character is $e^{-i\mu\phi}$. For a circle this is just Fourier analysis, for a general compact semisimple Lie group Weyl's generalization of Fourier analysis. Consider projection on the μ th subspace of the integral along the rotational direction

$$I_G = \int_G g \chi^{(\mu)}(g) \oint dx_G \delta_G(x(t)_G - (D(g)x(0))_G).$$

Coordinate x_G is the set of points swept by $[0, 2\pi]$ rotation of a point $x_0 = x_G(0, 0)$, so it is natural to parametrize it by the rotation angle ϕ' : $x_G = x(0, \phi')$, and rewrite the circle integral as

$$\oint dx_G \delta(x_G - D_\mu(g)x_G(t, 0)) = \int_0^{2\pi} d\phi' \frac{dx}{d\phi'}(0, \phi') \delta(x(0, \phi') - x(t, \phi' + \phi)).$$

Inverting the order of integrations,

$$I_G = \int_0^{2\pi} d\phi' \int_0^{2\pi} \frac{d\phi}{2\pi} e^{-i\mu\phi} \frac{dx}{d\phi}(0, \phi') \delta(x(0, \phi') - x(t, \phi' + \phi)) .$$

The integral is nonvanishing for smallest ϕ_p for which $x(0, \phi') = x(t, \phi' + \phi_p)$, and for all its repeats. Expand the argument of δ function in each such neighborhood $\phi' = \phi_p + \phi''$.

$$\begin{aligned} x(t, \phi' + \phi_p + \phi'') &= x(t, \phi' + \phi_p) + \phi'' \frac{dx}{d\phi}(t, \phi' + \phi_p) + \dots \\ &= x(t, \phi') + \phi'' \frac{dx}{d\phi}(t, \phi') + \dots . \end{aligned}$$

substituting back yields

$$\begin{aligned} I_G &= \int_0^{2\pi} \frac{d\phi'}{2\pi} \sum_{r=1}^{\infty} e^{-i\mu\phi_{p,r}} \frac{dx(0, \phi')}{d\phi} \int_{-\epsilon}^{\epsilon} d\phi'' e^{-i\mu\phi''} \delta(\phi'' \frac{dx}{d\phi}(0, \phi')) \\ &= \sum_{r=1}^{\infty} e^{-i\mu\phi_{p,r}} . \end{aligned}$$

[return to p. 532](#)

26.5 Exercises

- 26.1. **Haar measure for SU(2).** SU(2) acts on vectors in $\mathbb{C}^2$, and preserves their absolute value, hence its action can be parameterized by a 3-sphere S_3 , and multiplication can be viewed as an orthogonal transformation of S_3 . This is a special case of the formula $N = n^2 - 1$ for the dimension of SU(n). Show that the invariant Haar measure on SU(2)

$$\begin{aligned} & \int_{\text{SU}(2)} f(g) dg \\ &= \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_0^{\pi} \int_0^{\pi} \sin^2 \phi_1 \sin \phi_2 d\theta d\phi_1 d\phi_2 f(\theta, \phi_1, \phi_2) \end{aligned} \quad (26.34)$$

is a normalized surface measure on S_3 .

- 26.2. **Relative periodic orbits for circles, bagels and spheres:** (a) Show that relative periodic orbits for a point scattering specularly in a circular billiard are single scattering arcs. Compute their stability. Compute the spectrum.
- (b) Show that relative periodic orbits for a point scattering specularly in the plane that slices symmetrically upper half of a bagel (floating tire, torus) are single scattering arcs. Compute their stability. Compute the spectrum. Compute the escape rate.
- (c) Show that relative periodic orbits for a point scattering specularly within a sphere billiard are single scattering arcs. Compute their stability, spectrum.

Part III

Chaos: what to do about it?

WHAT YOU KNOW NOW is to partition topologically and invariantly the state space, compute a hierarchy of cycles, compute spectral determinants and their eigenvalues. What next?

1. Why cycle? (chapter 27)
2. Why does it work? (chapter 28)
3. When does it not work? (chapter 29)
4. What does it have to do with turbulence? (chapter 30)
5. How fat is the turbulent attractor? (chapter 30)
6. There is one that is experimentally relevant (chapter 32)

Chapter 27

Why cycle?

“Progress was a labyrinth ... people plunging blindly in and then rushing wildly back, shouting that they had found it ... the invisible king - the *élan vital* - the principle of evolution ... writing a book, starting a war, founding a school....”

—F. Scott Fitzgerald, *This Side of Paradise*

IN THE PRECEDING CHAPTERS we have moved rather briskly through the evolution operator formalism. Here we slow down in order to develop some fingertip feeling for the traces of evolution operators. It is a melancholy task, as the “intuition” garnered by these heuristic approximations is in all ways inferior to the straightforward and exact theory developed so far. But, it has to be done, as there is immense literature out there that deploys these heuristic estimates, most of it uninspired, some of it plain wrong, and the reader should be able to understand and sort through that literature. We start out by explaining qualitatively how local exponential instability of topologically distinct trajectories leads to a global exponential instability.

27.1 Escape rates

We start by verifying the claim (20.10) that for a nice hyperbolic flow the trace of the evolution operator grows exponentially with time. Consider again the game of pinball in figure 1.1. Designate by $\mathcal{M}$ a region of state space that encloses the three disks, such as the surface of the table along with all pinball directions. The fraction of initial points whose trajectories start within $\mathcal{M}$ and recur within that region at the time t is given by

$$\hat{\Gamma}_{\mathcal{M}}(t) = \frac{1}{|\mathcal{M}|} \int \int_{\mathcal{M}} dx dy \delta(y - f^t(x)) . \quad (27.1)$$

This quantity is both measurable and physically interesting in a variety of problems spanning nuclear physics to celestial mechanics. The integral over x takes

care of all possible initial pinballs; the integral over y checks whether they are still within $\mathcal{M}$ by time t . If the dynamics is bounded, and $\mathcal{M}$ envelops the entire accessible state space, $\hat{\Gamma}_{\mathcal{M}}(t) = 1$ for all t . However, if trajectories exit $\mathcal{M}$, the recurrence fraction decreases with time. For example, any trajectory that falls off the pinball table in figure 1.1 is gone for good.

These observations can be made more concrete by examining the pinball phase-space of figure 1.8. With each pinball bounce the initial conditions that survive get thinned out, each strip yielding two thinner strips within it. The total fraction of survivors (1.2) after n bounces is given by

$$\hat{\Gamma}_n = \frac{1}{|\mathcal{M}|} \sum_i^{(n)} |\mathcal{M}_i|, \quad (27.2)$$

where i is a binary label of the i th strip, and $|\mathcal{M}_i|$ is the area of the i th strip. Phase-space volume is preserved by the flow, so the strips of survivors are contracted along the stable eigen-directions and ejected along the unstable eigen-directions. As a crude estimate of the number of survivors in the i th strip, assume that a ray of trajectories spreads by a factor Λ after every bounce. The quantity Λ represents the mean value of the expanding eigenvalue of the corresponding Jacobian matrix of the flow. We replace $|\mathcal{M}_i|$ by the phase-space strip width estimate $|\mathcal{M}_i|/|\mathcal{M}| \sim 1/\Lambda_i$, which is right in spirit but not without drawbacks. For example, in general the eigenvalues of a Jacobian matrix for a finite segment of a trajectory have no invariant meaning; they depend on the choice of coordinates. However, we saw in chapter 21 that neighborhood sizes are determined by Floquet multipliers of periodic points, which are invariant under smooth coordinate transformations.

In the approximation $\hat{\Gamma}_n$ receives 2^n contributions of equal size

$$\hat{\Gamma}_1 \sim \frac{1}{\Lambda} + \frac{1}{\Lambda}, \dots, \hat{\Gamma}_n \sim \frac{2^n}{\Lambda^n} = e^{-n(\lambda-h)} = e^{-n\gamma}, \quad (27.3)$$

up to pre-exponential factors. We see here the interplay of the two key ingredients of chaos first mentioned in sect. 1.3.1: the escape rate γ equals the local expansion rate (the Lyapunov exponent $\lambda = \ln \Lambda$) minus the rate of global reinjection back into the system (the topological entropy $h = \ln 2$).

At each bounce one routinely loses the same fraction of trajectories, so one expects the sum (27.2) to decay exponentially with n . More precisely, by the hyperbolicity assumption of sect. 21.1.1, the expanding eigenvalue of the Jacobian matrix of the flow is exponentially bounded from both above and below,

$$1 < |\Lambda_{\min}| \leq |\Lambda(x)| \leq |\Lambda_{\max}|, \quad (27.4)$$

and the area of each strip in (27.2) is bounded by $|\Lambda_{\max}^{-n}| \leq |\mathcal{M}_i| \leq |\Lambda_{\min}^{-n}|$. Replacing $|\mathcal{M}_i|$ in (27.2) by its estimates in terms of $|\Lambda_{\max}|$ and $|\Lambda_{\min}|$ immediately leads to exponential bounds $(2/|\Lambda_{\max}|)^n \leq \hat{\Gamma}_n \leq (2/|\Lambda_{\min}|)^n$, i.e.,

$$\ln |\Lambda_{\max}| - \ln 2 \geq -\frac{1}{n} \ln \hat{\Gamma}_n \geq \ln |\Lambda_{\min}| - \ln 2. \quad (27.5)$$

The argument based on (27.5) establishes only that the sequence $\gamma_n = -\frac{1}{n} \ln \Gamma_n$ has a lower and an upper bound for any n . In order to prove that γ_n converge to the limit γ , we first show that for hyperbolic systems the sum over surviving intervals (27.2) can be replaced by a sum over periodic orbit stabilities. By (27.4) the size of the strip $\mathcal{M}_i$ can be bounded by the stability Λ_i of the i th periodic point:

$$C_1 \frac{1}{|\Lambda_i|} < \frac{|\mathcal{M}_i|}{|\mathcal{M}|} < C_2 \frac{1}{|\Lambda_i|}, \quad (27.6)$$

for any periodic point i of period n , with constants C_j dependent on the dynamical system but independent of n . The meaning of these bounds is that for increasingly long cycles in a system of bounded hyperbolicity, the shrinking of the i th strip is better approximated by the derivatives evaluated on the periodic point within the strip. Hence, the survival probability can be bounded close to the periodic point stability sum

$$\hat{C}_1 \Gamma_n < \sum_i^{(n)} \frac{|\mathcal{M}_i|}{|\mathcal{M}|} < \hat{C}_2 \Gamma_n, \quad (27.7)$$

where $\Gamma_n = \sum_i^{(n)} 1/|\Lambda_i|$ is the asymptotic trace sum (21.22). This establishes that for hyperbolic systems the survival probability sum (27.2) can be replaced by the periodic orbit sum (21.22).

exercise 27.1
exercise 19.4

We conclude that for hyperbolic, locally unstable flows the fraction (27.1) of initial x whose trajectories remain trapped within $\mathcal{M}$ up to time t is expected to decay exponentially,

$$\Gamma_{\mathcal{M}}(t) \propto e^{-\gamma t},$$

where γ is the asymptotic *escape rate* defined by

$$\gamma = - \lim_{t \rightarrow \infty} \frac{1}{t} \ln \Gamma_{\mathcal{M}}(t). \quad (27.8)$$

27.2 Natural measure in terms of periodic orbits

Let us now refine the reasoning of sect. 27.1 and argue that the trace is a discretized integral over state space. Consider the trace (21.6) in the large time limit (21.21):

$$\text{tr } \mathcal{L}^n = \int dx \delta(x - f^n(x)) e^{\beta A(x,n)} \approx \sum_i^{(n)} \frac{e^{\beta A(x_i,n)}}{|\Lambda_i|}.$$

The factor $1/|\Lambda_i|$ was interpreted in (27.2) as the area of the i th phase-space strip. Hence, the $\text{tr } \mathcal{L}^n$ represents a discrete version of $\int dx e^{\beta A(x,n)}$ approximated by a tessellation into strips centered on periodic points x_i , (see figure 1.10), with the volume of the i th neighborhood given by estimate $|\mathcal{M}_i| \sim 1/|\Lambda_i|$, and $e^{\beta A(x,n)}$ estimated by $e^{\beta A(x_i,n)}$, its value at the i th periodic point. If the symbolic dynamics is complete, any state space rectangle $[s_{-m} \cdots s_0.s_1 s_2 \cdots s_n]$ always contains the periodic point $\overline{s_{-m} \cdots s_0 s_1 s_2 \cdots s_n}$; hence, although the periodic points are of

section 15.3.1

measure zero (just like rationals in the unit interval), they are dense on the non-wandering set. Equipped with a measure for the associated rectangles, periodic orbits suffice to cover the entire non-wandering set. The average of $e^{\beta A}$ evaluated on the non-wandering set is therefore given by the trace, properly normalized so that $\langle 1 \rangle = 1$:

$$\langle e^{\beta A} \rangle_n \approx \frac{\sum_i^{(n)} e^{\beta A(x_i, n)} / |\Lambda_i|}{\sum_i^{(n)} 1 / |\Lambda_i|} = \sum_i^{(n)} \mu_i e^{\beta A(x_i, n)}. \quad (27.9)$$

Here μ_i is the *normalized natural measure*

section 20.3

$$\sum_i^{(n)} \mu_i = 1, \quad \mu_i = e^{n\gamma_n} / |\Lambda_i|, \quad (27.10)$$

which is correct both for closed systems as well as open systems.

Unlike brute numerical slicing of the integration space into an arbitrary lattice (for a critique, see sect. 19.3), periodic orbit theory is smart, as it automatically partitions integrals according to the intrinsic topology of the flow, and assigns to each tile i the invariant natural measure μ_i .

27.2.1 Unstable periodic orbits are dense

(L. Rondoni and P. Cvitanović)

Our goal in sect. 20.1 was to evaluate the space and time averaged expectation value (20.8). An average over all periodic orbits can accomplish the job only if the periodic orbits fully explore the asymptotically accessible state space.

Why should unstable periodic points end up being dense? The cycles are intuitively expected to be *dense* because on a connected chaotic set a typical trajectory is expected to behave ergodically, and infinitely many times pass arbitrarily close to any point on the set (including the initial point of the trajectory itself). The argument proceeds more or less as follows. Partition $\mathcal{M}$ in arbitrarily small regions and consider particles that start in the region $\mathcal{M}_i$, and return to it in n steps after some peregrination in the state space. For example, a trajectory might return a little to the left of its original position, whereas a nearby neighbor might return a little to the right of its original position. By assumption, the flow is continuous, so generically one expects to be able to gently move the initial point in such a way that the trajectory returns precisely to the initial point, i.e., one expects a periodic point of period n in cell i . As we diminish the size of regions $\mathcal{M}_i$, aiming a trajectory that returns to $\mathcal{M}_i$ becomes increasingly difficult. Therefore, we are guaranteed that unstable orbits of increasingly large periods are densely interspersed in the asymptotic non-wandering set.

The above argument is heuristic, by no means guaranteed to work, and it must be checked for the particular system at hand. A variety of ergodic but insufficiently mixing counter-examples can be constructed - the most familiar being a quasiperiodic motion on a torus.

27.3 Correlation functions

The *time correlation function* $C_{AB}(t)$ of two observables A and B along the trajectory $x(t) = f^t(x_0)$ is defined as

$$C_{AB}(t; x_0) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T d\tau A(x(\tau + t))B(x(\tau)), \quad x_0 = x(0). \quad (27.11)$$

If the system is ergodic, with invariant continuous measure $\rho_0(x)dx$, then correlation functions do not depend on x_0 (apart from a set of zero measure), and may be computed by a state space average as well,

$$C_{AB}(t) = \int_{\mathcal{M}} dx_0 \rho_0(x_0) A(f^t(x_0)) B(x_0). \quad (27.12)$$

For a chaotic system we expect that time evolution will lose the information contained in the initial conditions, so that $C_{AB}(t)$ will approach the *uncorrelated* limit $\langle A \rangle \cdot \langle B \rangle$. As a matter of fact the asymptotic decay of correlation functions

$$\hat{C}_{AB} := C_{AB} - \langle A \rangle \langle B \rangle \quad (27.13)$$

for any pair of observables coincides with the definition of *mixing*, a fundamental property in ergodic theory. We now assume without loss of generality that $\langle B \rangle = 0$. (Otherwise we may define a new observable by $B(x) - \langle B \rangle$.) Our purpose is now to connect the asymptotic behavior of correlation functions with the spectrum of the Perron-Frobenius operator $\mathcal{L}$. We can write (27.12) as

$$\tilde{C}_{AB}(t) = \int_{\mathcal{M}} dx \int_{\mathcal{M}} dy A(y) B(x) \rho_0(x) \delta(y - f^t(x))$$

and recover the evolution operator

$$\tilde{C}_{AB}(t) = \int_{\mathcal{M}} dx \int_{\mathcal{M}} dy A(y) \mathcal{L}^t(y, x) B(x) \rho_0(x).$$

Recall sect. 19.1, where we showed that $\rho(x)$ is the eigenvector of $\mathcal{L}$ corresponding to probability conservation:

$$\int_{\mathcal{M}} dy \mathcal{L}^t(x, y) \rho(y) = \rho(x).$$

We can expand the x -dependent part of this equation in terms of the eigenbasis of $\mathcal{L}$:

$$B(x) \rho_0(x) = \sum_{\alpha=0}^{\infty} c_{\alpha} \rho_{\alpha}(x),$$

where $\rho_0(x)$ is the natural measure. Since the average of the left hand side is zero the coefficient c_0 must vanish. The action of $\mathcal{L}$ can then be written as

$$\tilde{C}_{AB}(t) = \sum_{\alpha \neq 0} e^{-s_{\alpha} t} c_{\alpha} \int_{\mathcal{M}} dy A(y) \rho_{\alpha}(y). \quad (27.14)$$

We see immediately that if the spectrum has a *gap*, i.e., if the second largest leading eigenvalue is isolated from the largest eigenvalue ($s_0 = 0$) then (27.14) implies *exponential* decay of correlations

exercise 27.2

$$\tilde{C}_{AB}(t) \sim e^{-\nu t}.$$

The correlation decay rate $\nu = s_1$ then depends only on intrinsic properties of the dynamical system (the position of the next-to-leading eigenvalue of the Perron-Frobenius operator), and the choice of a particular observable influences only the prefactor.

Correlation functions are often accessible from time series measurable in laboratory experiments and numerical simulations; moreover, they are intimately linked to transport exponents.

27.4 Trace formulas vs. level sums

Benoit B. Mandelbrot: “I would be perfectly happy being Kepler” [to a coming fractals’ Newton]. Referring to the broad array of things now described by fractals, he added, “I have been Kepler many times over.”

—J. Gleick, New York Times, January 22, 1985



Trace formulas (21.9) and (21.19) diverge precisely where one would like to use them, at s equal to eigenvalues s_α . To avoid this divergence, one can proceed as follows; according to (21.23) the “level” sums (all symbol strings of length n) are asymptotically dominated by the leading eigenvalue $e^{s_0 n}$ of the evolution operator

$$\sum_{x_i \in \text{Fix}_f^n} \frac{e^{\beta A(x_i, n)}}{|\Lambda_i|} \rightarrow e^{s_0 n},$$

so an n th order estimate $s_{(n)}$ of the leading eigenvalue s_0 is fixed by the condition

$$1 = \sum_{x_i \in \text{Fix}_f^n} \frac{e^{\beta A(x_i, n)} e^{-s_{(n)} n}}{|\Lambda_i|}. \quad (27.15)$$

The eigenvalue condition for the level sum (27.15) can be written in the same form as the two conditions (23.18) and (23.19) given so far:

$$0 = 1 - \sum_i^{(n)} t_i, \quad t_i = t_i(\beta, s(\beta)), \quad n_i = n. \quad (27.16)$$

We do not recommend it as a computational method. The difficulty in estimating the leading eigenvalue s_0 from this $n \rightarrow \infty$ limit is at least twofold:

1. Due to an exponential growth in the number of intervals and an exponential decrease in the attainable accuracy, the maximum n , achieved experimentally or numerically, is approximately between 5 and 20.

2. The pre-asymptotic sequence of finite estimates $s_{(n)}$ is not unique, because the sums Γ_n depend on how we define the escape region, and because in general the areas $|\mathcal{M}_i|$ in the sum (27.2) should be weighted by the density of initial conditions $\rho(0)$. For example, an overall measuring unit rescaling $|\mathcal{M}_i| \rightarrow \alpha|\mathcal{M}_i|$ introduces $1/n$ corrections in $s_{(n)}$ defined by the log of the sum (27.8): $s_{(n)} \rightarrow s_{(n)} + \ln \alpha/n$. This problem can be ameliorated by defining a level average as a function of s ,

$$\langle e^{\beta A(s)} \rangle_{(n)} := \sum_{x_i \in \text{Fix } f^n} \frac{e^{\beta A(x_i, n)} e^{sn}}{|\Lambda_i|}, \quad (27.17)$$

and determining the n th level estimate $s_{(n)}$ by requiring that the ratios of successive levels satisfy

$$1 = \frac{\langle e^{\beta A(s_{(n)})} \rangle_{(n)}}{\langle e^{\beta A(s_{(n)})} \rangle_{(n-1)}}.$$

This avoids the worst problem with formula (27.15), the $1/n$ corrections due to its lack of rescaling invariance. However, even though much published pondering of “chaos” relies on it, there is no need for such gymnastics: dynamical zeta functions and spectral determinants are already invariant not only under linear rescalings, but under *all* smooth nonlinear conjugacies $x \rightarrow h(x)$, and require no $n \rightarrow \infty$ extrapolations to asymptotic times. Comparing this with cycle expansions (23.8), we see the difference; in the level sum approach, we keep increasing exponentially the number of terms using no information of what is already known from shorter estimates, but in cycle expansions short terms dominate and longer ones enter only as exponentially small corrections.

27.4.1 Flow conservation sum rules

The trace formula version of the flow conservation sum rule (23.17) comes in two varieties (one for maps and another for flows). By flow conservation, the leading eigenvalue is $s_0 = 0$, which for maps (27.16) yields

$$\text{tr } \mathcal{L}^n = \sum_{x_i \in \text{Fix } f^n} \frac{1}{|\det(\mathbf{1} - M^n(x_i))|} = 1 + e^{s_1 n} + \dots \quad (27.18)$$

For flows, one can apply this rule by grouping together cycles from $t = T$ to $t = T + \Delta T$

$$\begin{aligned} \frac{1}{\Delta T} \sum_{p, r}^{T \leq r T_p \leq T + \Delta T} \frac{T_p}{|\det(\mathbf{1} - M_p^r)|} &= \frac{1}{\Delta T} \int_T^{T + \Delta T} dt (1 + e^{s_1 t} + \dots) \\ &= 1 + \frac{1}{\Delta T} \sum_{\alpha=1}^{\infty} \frac{e^{s_\alpha T}}{s_\alpha} (e^{s_\alpha \Delta T} - 1) \approx 1 + e^{s_1 T} + \dots \end{aligned} \quad (27.19)$$

As is usual for fixed level trace sums, the convergence of (27.18) is controlled by the gap between the leading and next-to-leading eigenvalues of the evolution operator.

Résumé

We conclude this chapter by a general comment on the relation of finite trace sums such as (27.2) to spectral determinants and dynamical zeta functions. One might be tempted to believe that given a deterministic rule, a sum like (27.2) can be evaluated to any desired precision. For short times, this is indeed true: every region M_i in (27.2) can be accurately delineated, and there is no need for any fancy theory. However, if the dynamics is unstable, local variations in initial conditions grow exponentially and in a finite time attain the size of the system. The difficulty with estimating the $n \rightarrow \infty$ limit from (27.2) is then at least twofold:

1. Due to the exponential growth in number of intervals, and the exponential decrease in attainable accuracy, the maximal n attainable experimentally or numerically is in practice of order of something between 5 to 20;

2. The pre-asymptotic sequence of finite estimates γ_n is not unique, because the sums $\hat{\Gamma}_n$ depend on how we define the escape region, and because in general the areas $|M_i|$ in the sum (27.2) should be weighted by the density of initial x_0 .

In contrast, dynamical zeta functions and spectral determinants are invariant under *all* smooth nonlinear conjugacies $x \rightarrow h(x)$, not only linear rescalings, and require no $n \rightarrow \infty$ extrapolations.

Commentary

Remark 27.1. Nonhyperbolic measures. The measure $\mu_i = 1/|\Lambda_i|$ is the natural measure only for the strictly hyperbolic systems. For nonhyperbolic systems, the measure might develop cusps. For example, for Ulam maps (unimodal maps with quadratic critical point mapped onto the “left” unstable fixed point x_0 , discussed in more detail in chapter 29), the measure develops a square-root singularity on the $\bar{0}$ cycle:

$$\mu_0 = \frac{1}{|\Lambda_0|^{1/2}} . \quad (27.20)$$

Thermodynamic averages are still expected to converge in the “hyperbolic” phase in which the positive entropy of unstable orbits dominates the marginal orbits, but they fail in the “nonhyperbolic” phase. The general case remains unclear [5, 7, 8, 10, 13].

Remark 27.2. Trace formula periodic orbit averaging. The cycle averaging formulas are not the first thing one intuitively writes down; the approximate trace formulas are more accessibly heuristically. Trace formula for averaging (27.19) was discussed for the first time by Hannay and Ozorio de Almeida [11, 12]. Another novelty of cycle averaging formulas is one of their main virtues, in contrast to the explicit analytical results such as those of ref. [9]. Their evaluation *does not* require any explicit construction of the (coordinate dependent) eigenfunctions of the Perron-Frobenius operator (i.e., the natural measure ρ_0).

Remark 27.3. Role of noise in dynamical systems. In any physical application, the dynamics is always accompanied by external noise in addition to deterministic chaos.

The former can be characterized by its strength σ and distribution. Lyapunov exponents, correlation decay, and dynamo rate can be defined in this case the same way as in the deterministic case. One might think that noise completely destroys the results derived here. However, as we show chapter 33, deterministic formulas remain valid to accuracy comparable with noise width if the noise level is small. A small level of noise even helps, as it makes the dynamics more ergodic. Deterministically non-communicating parts of state space become weakly connected due to noise. This argument explains why periodic orbit theory is also applicable to non-ergodic systems. For small amplitude noise, one can expand perturbatively

$$\bar{a} = \bar{a}_0 + \bar{a}_1 \sigma^2 + \bar{a}_2 \sigma^4 + \dots,$$

around the deterministic averages a_0 . The expansion coefficients $\bar{a}_1, \bar{a}_2, \dots$ can also be expressed in terms of periodic orbit formulas. Calculating these coefficients is one of the challenges facing periodic orbit theory, discussed in refs. [3, 4, 6].

References

- [1] R. Artuso, E. Aurell, and P. Cvitanović, “Recycling of strange sets: II. Applications”, *Nonlinearity* **3**, 361–386 (1990).
- [2] F. Christiansen, G. Paladin, and H. H. Rugh, “Determination of correlation spectra in chaotic systems”, *Phys. Rev. Lett.* **65**, 2087–2090 (1990).
- [3] P. Cvitanović, C. P. Dettmann, R. Mainieri, and G. Vattay, “Trace formulas for stochastic evolution operators: Weak noise perturbation theory”, *J. Stat. Phys.* **93**, 981–999 (1998).
- [4] P. Cvitanović, C. P. Dettmann, R. Mainieri, and G. Vattay, “Trace formulae for stochastic evolution operators: Smooth conjugation method”, *Nonlinearity* **12**, 939 (1999).
- [5] P. Cvitanović, G. H. Gunaratne, and I. Procaccia, “Topological and metric properties of Hénon-type strange attractors”, *Phys. Rev. A* **38**, 1503–1520 (1988).
- [6] P. Cvitanović, N. Sørensgaard, G. Palla, G. Vattay, and C. P. Dettmann, “Spectrum of stochastic evolution operators: Local matrix representation approach”, *Phys. Rev. E* **60**, 3936–3941 (1999).
- [7] P. Grassberger, R. Badii, and A. Politi, “Scaling laws for invariant measures on hyperbolic and nonhyperbolic attractors”, *J. Stat. Phys.* **51**, 135–178 (1988).
- [8] C. Grebogi, E. Ott, and J. A. Yorke, “Crises, sudden changes in chaotic attractors, and transient chaos”, *Physica D* **7**, 181–200 (1983).
- [9] S. Grossmann and S. Thomae, “Invariant distributions and stationary correlation functions of one-dimensional discrete processes”, *Z. Naturf. A* **32**, 1353–1363 (1977).
- [10] E. Ott, C. Grebogi, and J. A. Yorke, “Theory of first order phase transitions for chaotic attractors of nonlinear dynamical systems”, *Phys. Lett. A* **135**, 343–348 (1989).

- [11] A. M. Ozorio de Almeida, *Hamiltonian Systems: Chaos and Quantization* (Cambridge Univ. Press, Cambridge UK, 1989).
- [12] A. M. Ozorio de Almeida and J. H. Hannay, “Periodic orbits and a correlation function for the semiclassical density of states”, *J. Phys. A* **17**, 3429 (1984).
- [13] A. Politi, R. Badii, and P. Grassberger, “On the geometric structure of non-hyperbolic attractors”, *J. Phys. A* **21**, L763–L769 (1988).

27.5 Exercises

27.1. Escape rate of the logistic map.

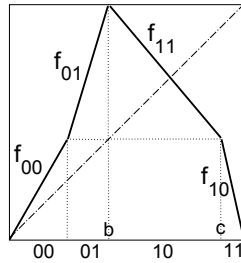
- (a) Calculate the fraction of trajectories remaining trapped in the interval $[0, 1]$ for the logistic map

$$f(x) = A(1 - (2x - 1)^2), \quad (27.21)$$

and determine the A dependence of the escape rate $\gamma(A)$ numerically.

- (b) Develop a numerical method for calculating the lengths of intervals of trajectories remaining stuck for n iterations of the map.
- (c) Describe the dependence of A near the critical value $A_c = 1$?

27.2. Four-scale map correlation decay rate. Consider the piecewise-linear map



$$f(x) = \begin{cases} f_{00} = \Lambda_0 x \\ f_{01} = s_{01}(x - b) + 1 \\ f_{11} = \Lambda_1(x - b) + 1 \\ f_{10} = s_{10}(x - 1) \end{cases}$$

with a 4-interval state space Markov partition

$$\begin{aligned} \mathcal{M} &= \{\mathcal{M}_{00}, \mathcal{M}_{01}, \mathcal{M}_{10}, \mathcal{M}_{11}\} \\ &= \{[0, b/\Lambda_0], (b/\Lambda_0, b], (b, c], (c, 1]\}. \end{aligned}$$

- (a) compute s_{01} , s_{10} , c .
- (b) Show that the 2-cycle Floquet multiplier does not depend on b ,

$$\Lambda_{01} = s_{01}s_{10} = -\frac{\Lambda_0\Lambda_1}{(\Lambda_0 - 1)(\Lambda_1 + 1)}.$$

- (c) Write down the $[2 \times 2]$ Perron-Frobenius operator acting on the space of densities piecewise constant over the four partitions.
- (d) Construct the corresponding transition graph.
- (e) Write down the corresponding spectral determinant.
- (f) Show that the escape rate vanishes, $\gamma = -\ln(z_0) = 0$.
- (g) Determine the spectrum of the Perron-Frobenius operator on the space of densities piecewise constant over the four partitions. Show that the second largest eigenvalue of the is $\frac{1}{z_1} = -1 + \frac{1}{\Lambda_0} - \frac{1}{\Lambda_1}$.
- (h) Is this value consistent with the tent map value previously computed in exercise 19.4 (with the appropriate choice of $\{\Lambda_0, \Lambda_1, c\}$).

- (i) (optional) Is this next-to leading eigenvalue still correct if the Perron-Frobenius operator acts on the space of analytic functions?

27.3. Lyapunov exponents for 1-dimensional maps. Extend your cycle expansion programs so that the first and the second moments of observables can be computed. Use it to compute the Lyapunov exponent for the following maps:

- (a) the piecewise-linear skew tent (flow conserving map)

$$f(x) = \begin{cases} \Lambda_0 x & \text{if } 0 \leq x < \Lambda_0^{-1}, \\ \Lambda_1(1-x) & \text{if } \Lambda_0^{-1} \leq x \leq 1. \end{cases},$$

$$\Lambda_1 = \Lambda_0/(\Lambda_0 - 1).$$

- (b) the Ulam map $f(x) = 4x(1-x)$.

- (c) the skew Ulam map

$$f(x) = \Lambda_0 x(1-x)(1-bx), \quad (27.22)$$

$1/\Lambda_0 = x_c(1-x_c)(1-bx_c)$. In our numerical work we fix (arbitrarily, the value chosen in ref. [1]) $b = 0.6$, so

$$f(x) = 0.1218 x(1-x)(1-0.6x)$$

with a peak $f(x_c) = 1$ at $x_c = 0.7$.

- (d) the repeller of $f(x) = Ax(1-x)$, for either $A = 9/2$ or $A = 6$ (this is a continuation of exercise 23.2).

- (e) the 2-branch flow conserving map

$$\begin{aligned} f_0(x) &= \frac{1}{2h} \left(h - p + \sqrt{(h-p)^2 + 4hx} \right) \\ f_1(x) &= \frac{1}{2h} (h + p - 1) \\ &\quad + \frac{1}{2h} \sqrt{(h + p - 1)^2 + 4h(x - p)}, \end{aligned} \quad (27.23)$$

with a 2-interval state space Markov partition $\mathcal{M} = \{\mathcal{M}_0, \mathcal{M}_1\} = \{[0, p], (p, 1]\}$. This is a nonlinear perturbation of the Bernoulli shift map, for which $h = 0$ (28.19); the first 15 eigenvalues of the Perron-Frobenius operator are listed in ref. [2] for $p = 0.8$, $h = 0.1$. Use these parameter values when computing the Lyapunov exponent.

Cases (a) and (b) can be computed analytically; cases (c), (d) and (e) require numerical computation of cycle stabilities. Just to see whether the theory is worth the trouble, also check your cycle expansions results for cases (c) and (d) with Lyapunov exponents computed by direct numerical averaging along trajectories of randomly chosen initial points:

- (f) trajectory-trajectory separation (6.1) (hint: rescale δx every so often, to avoid numerical overflows),
 (g) iterated stability (6.12).

How good is the numerical accuracy compared with periodic orbit theory predictions for (a) - (g)?

Chapter 28

Why does it work?

Bloch: “Space is the field of linear operators.”

Heisenberg: “Nonsense, space is blue and birds fly through it.”

—Felix Bloch, *Heisenberg and the early days of quantum mechanics*

(R. Artuso, H.H. Rugh and P. Cvitanović)

AS WE SHALL SEE, the trace formulas and spectral determinants work well, sometimes very well. The question is: Why? And it still is. The heuristic manipulations of chapter 21 were naive and reckless, as we are facing infinite-dimensional vector spaces and singular integral kernels.

We now outline the key ingredients of proofs that put the trace and determinant formulas on solid footing. This requires taking a closer look at the evolution operators from a mathematical point of view, since up to now we have talked about eigenvalues without any reference to what kind of a function space the corresponding eigenfunctions belong to. We shall restrict our considerations to the spectral properties of the Perron-Frobenius operator for maps, as proofs for more general evolution operators follow along the same lines. What we refer to as a “the set of eigenvalues” acquires meaning only within a precisely specified functional setting: this sets the stage for a discussion of the analyticity properties of spectral determinants. In example 28.1 we compute explicitly the eigenspectrum for the three analytically tractable piecewise linear examples. In sect. 28.3 we review the basic facts of the classical Fredholm theory of integral equations. The program is sketched in sect. 28.4, motivated by an explicit study of eigenspectrum of the Bernoulli shift map, and in sect. 28.5 generalized to piecewise real-analytic hyperbolic maps acting on appropriate densities. We show on a very simple example that the spectrum is quite sensitive to the regularity properties of the functions considered.

For expanding and hyperbolic finite-subshift maps analyticity leads to a very strong result; not only do the determinants have better analyticity properties than

the trace formulas, but the spectral determinants are singled out as entire functions in the complex s plane.

remark 28.1

The goal of this chapter is not to provide an exhaustive review of the rigorous theory of the Perron-Frobenius operators and their spectral determinants, but rather to give you a feeling for how our heuristic considerations can be put on a firm basis. The mathematics underpinning the theory is both hard and profound.

If you are primarily interested in applications of the periodic orbit theory, you should skip this chapter on the first reading.

28.1 Linear maps: exact spectra

We start gently; in example 28.1 we work out the *exact* eigenvalues and eigenfunctions of the Perron-Frobenius operator for the simplest example of unstable, expanding dynamics, a linear 1-dimensional map with one unstable fixed point. Ref. [12] shows that this can be carried over to d -dimensions. Not only that, but in example 28.2 we compute the exact spectrum for the simplest example of a dynamical system with an *infinity* of unstable periodic orbits, the Bernoulli shift.

 example 28.1
p. 575

What are these eigenfunctions? Think of eigenfunctions of the Schrödinger equation: k labels the k th eigenfunction x^k in the same spirit in which the number of nodes labels the k th quantum-mechanical eigenfunction. A quantum-mechanical amplitude with more nodes has more variability, hence a higher kinetic energy. Analogously, for a Perron-Frobenius operator, a higher k eigenvalue $1/|\Lambda|^k$ is getting exponentially smaller because densities that vary more rapidly decay more rapidly under the expanding action of the map.

 example 28.3
p. 576

The left hand side of (28.22) is a meromorphic function, with the leading zero at $z = |\Lambda|$. So what?

 example 28.4
p. 576

This example shows that: (1) an estimate of the leading pole (the leading eigenvalue of $\mathcal{L}$) from a finite truncation of a trace formula converges exponentially, and (2) the non-leading eigenvalues of $\mathcal{L}$ lie outside of the radius of convergence of the trace formula and cannot be computed by means of such cycle expansion. However, as we shall now see, the whole spectrum is reachable at no extra effort, by computing it from a determinant rather than a trace.

 example 28.5
p. 576

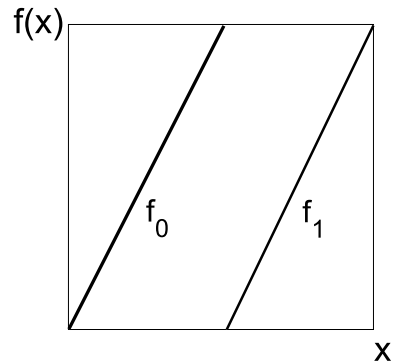


Figure 28.1: The Bernoulli shift map.

The main lesson to glean from this simple example is that the cumulants Q_n decay asymptotically *faster* than exponentially, as $\Lambda^{-n(n-1)/2}$. For example, if we approximate series such as (28.23) by the first 10 terms, the error in the estimate of the leading zero is $\approx 1/\Lambda^{50}$!

So far all is well for a rather boring example, a dynamical system with a single repelling fixed point. What about chaos? Systems where the number of unstable cycles increases exponentially with their period? We now turn to the simplest example of a dynamical system with an infinity of unstable periodic orbits.

The Bernoulli map spectrum looks reminiscent of the single fixed-point spectrum (28.18), with the difference that the leading eigenvalue here is 1, rather than $1/|\Lambda|$. The difference is significant: the single fixed-point map is a repeller, with escape rate (1.7) given by the $\mathcal{L}$ leading eigenvalue $\gamma = \ln |\Lambda|$, while there is no escape in the case of the Bernoulli map. As already noted in discussion of the relation (22.15), for bounded systems the local expansion rate (here $\ln |\Lambda| = \ln 2$) is balanced by the entropy (here $\ln 2$, the log of the number of preimages F_s), yielding zero escape rate.

section 22.4



example 28.2
p. 575

So far we have demonstrated that our periodic orbit formulas are correct for two piecewise linear maps in 1 dimension, one with a single fixed point, and one with a full binary shift chaotic dynamics. For a single fixed point, eigenfunctions are monomials in x . For the chaotic example, they are orthogonal polynomials on the unit interval. What about higher dimensions? We check our formulas on a 2-dimensional hyperbolic map next.



example 28.6
p. 576

So far we have checked the trace and spectral determinant formulas derived heuristically in chapters 21 and 22, but only for the case of 1-dimensional and 2-dimensional linear maps. But for infinite-dimensional vector spaces this game is fraught with dangers, and we have already been misled by piecewise linear examples into spectral confusions: contrast the spectra of example 19.1 and example 20.4 with the spectrum computed in example 21.2.

We show next that the above results do carry over to a sizable class of piecewise analytic expanding maps.

28.2 Evolution operator in a matrix representation

The standard, and for numerical purposes sometimes very effective way to look at operators is through their matrix representations. Evolution operators are moving density functions defined over some state space, and as in general we can implement this only numerically, the temptation is to discretize the state space as in sect. 19.3. The problem with such state space discretization approaches is that they sometimes yield plainly wrong spectra (compare example 20.4 with the result of example 21.2), so we have to think through carefully what is it that we *really* measure.

An expanding map $f(x)$ takes an initial smooth density $\phi_n(x)$, defined on a subinterval, stretches it out and overlays it over a larger interval, resulting in a new, smoother density $\phi_{n+1}(x)$. Repetition of this process smoothes the initial density, so it is natural to represent densities $\phi_n(x)$ by their Taylor series. Expanding

$$\phi_n(y) = \sum_{k=0}^{\infty} \phi_n^{(k)}(0) \frac{y^k}{k!}, \quad \phi_{n+1}(y)_k = \sum_{\ell=0}^{\infty} \phi_{n+1}^{(\ell)}(0) \frac{y^\ell}{\ell!},$$

$$\phi_{n+1}^{(\ell)}(0) = \int dx \delta^{(\ell)}(y - f(x)) \phi_n(x) \Big|_{y=0}, \quad x = f^{-1}(0),$$

and substitute the two Taylor series into (19.6):

$$\phi_{n+1}(y) = (\mathcal{L}\phi_n)(y) = \int_{\mathcal{M}} dx \delta(y - f(x)) \phi_n(x).$$

The matrix elements follow by evaluating the integral

$$\mathbf{L}_{\ell k} = \frac{\partial^\ell}{\partial y^\ell} \int dx \mathcal{L}(y, x) \frac{x^k}{k!} \Big|_{y=0}. \quad (28.1)$$

we obtain a matrix representation of the evolution operator

$$\int dx \mathcal{L}(y, x) \frac{x^k}{k!} = \sum_{k'} \frac{y^{k'}}{k'!} \mathbf{L}_{k'k}, \quad k, k' = 0, 1, 2, \dots$$

which maps the x^k component of the density of trajectories $\phi_n(x)$ into the $y^{k'}$ component of the density $\phi_{n+1}(y)$ one time step later, with $y = f(x)$.

We already have some practice with evaluating derivatives $\delta^{(\ell)}(y) = \frac{\partial^\ell}{\partial y^\ell} \delta(y)$ from sect. 19.2. This yields a representation of the evolution operator centered on the fixed point, evaluated recursively in terms of derivatives of the map f :

$$\begin{aligned} \mathbf{L}_{\ell k} &= \int dx \delta^{(\ell)}(x - f(x)) \frac{x^k}{k!} \Big|_{x=f(x)} \\ &= \frac{1}{|f'|} \left(\frac{d}{dx} \frac{1}{f'(x)} \right)^\ell \frac{x^k}{k!} \Big|_{x=f(x)}. \end{aligned} \quad (28.2)$$

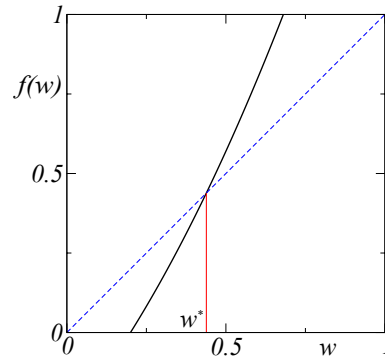


Figure 28.2: A nonlinear one-branch repeller with a single fixed point w_q .

The matrix elements vanish for $\ell < k$, so $\mathbf{L}$ is a lower triangular matrix. The diagonal and the successive off-diagonal matrix elements are easily evaluated iteratively by computer algebra

$$\mathbf{L}_{kk} = \frac{1}{|\Lambda|\Lambda^k}, \quad \mathbf{L}_{k+1,k} = -\frac{(k+2)!f''}{2k!|\Lambda|\Lambda^{k+2}}, \quad \dots$$

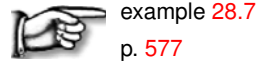
For chaotic systems the map is expanding, $|\Lambda| > 1$. Hence the diagonal terms drop off exponentially, as $1/|\Lambda|^{k+1}$, the terms below the diagonal fall off even faster, and truncating $\mathbf{L}$ to a finite matrix introduces only exponentially small errors.

The trace formula (28.22) takes now a matrix form

$$\text{tr} \frac{z\mathcal{L}}{1-z\mathcal{L}} = \text{tr} \frac{z\mathbf{L}}{1-z\mathbf{L}}. \quad (28.3)$$

In order to illustrate how this works, we work out a few examples.

In example 28.7 we show that these results carry over to any analytic single-branch 1-dimensional repeller. Further examples motivate the steps that lead to a proof that spectral determinants for general analytic 1-dimensional expanding maps, and - in sect. 28.5, for 1-dimensional hyperbolic mappings - are also entire functions.



example 28.7
p. 577

This super-exponential decay of cumulants Q_k ensures that for a repeller consisting of a single repelling point the spectral determinant (28.23) is *entire* in the complex z plane.

In retrospect, the matrix representation method for solving the density evolution problems is eminently sensible — after all, that is the way one solves a close relative to deterministic density evolution equations, the Schrödinger equation. When available, matrix representations for $\mathcal{L}$ enable us to compute many more orders of cumulant expansions of spectral determinants and many more eigenvalues of evolution operators than the cycle expansions approach.

Now, if the spectral determinant is entire, formulas such as (22.24) imply that the dynamical zeta function is a meromorphic function. The practical import of

this observation is that it guarantees that finite order estimates of zeroes of dynamical zeta functions and spectral determinants converge exponentially, or - in cases such as (28.23) - super-exponentially to the exact values, and so the cycle expansions to be discussed in chapter 23 represent a *true perturbative* approach to chaotic dynamics.

Before turning to specifics we summarize a few facts about classical theory of integral equations, something you might prefer to skip on first reading. The purpose of this exercise is to understand that the Fredholm theory, a theory that works so well for the Hilbert spaces of quantum mechanics does not necessarily work for deterministic dynamics - the ergodic theory is much harder.



fast track:
sect. 28.4, p. 562

28.3 Deterministic Fredholm theory

He who would valiant be 'gainst all disaster
Let him in constancy follow the Master.
—John Bunyan, *Pilgrim's Progress*



The Perron-Frobenius operator

$$\mathcal{L}\phi(x) = \int dy \delta(x - f(y)) \phi(y)$$

has the same appearance as a deterministic Fredholm integral operator

$$\mathcal{K}\varphi(x) = \int_{\mathcal{M}} dy \mathcal{K}(x, y) \varphi(y), \quad (28.4)$$

and one is tempted to resort to deterministic Fredholm theory in order to establish analyticity properties of spectral determinants. This path to enlightenment is blocked by the singular nature of the kernel, which is a distribution, whereas the standard theory of integral equations usually concerns itself with regular kernels $\mathcal{K}(x, y) \in L^2(\mathcal{M}^2)$. Here we briefly recall some steps of Fredholm theory, before working out the example of example 28.2.

The general form of Fredholm integral equations of the second kind is

$$\varphi(x) = \int_{\mathcal{M}} dy \mathcal{K}(x, y) \varphi(y) + \xi(x) \quad (28.5)$$

where $\xi(x)$ is a given function in $L^2(\mathcal{M})$ and the kernel $\mathcal{K}(x, y) \in L^2(\mathcal{M}^2)$ (Hilbert-Schmidt condition). The natural object to study is then the linear integral operator (28.4), acting on the Hilbert space $L^2(\mathcal{M})$: the fundamental property that follows from the $L^2(Q)$ nature of the kernel is that such an operator is *compact*, that is

close to a finite rank operator. A compact operator has the property that for every $\delta > 0$ only a *finite* number of linearly independent eigenvectors exist corresponding to eigenvalues whose absolute value exceeds δ , so we immediately realize (figure 28.5) that much work is needed to bring Perron-Frobenius operators into this picture.

We rewrite (28.5) in the form

$$\mathcal{T}\varphi = \xi, \quad \mathcal{T} = \mathbb{1} - \mathcal{K}. \quad (28.6)$$

The Fredholm alternative is now applied to this situation as follows: the equation $\mathcal{T}\varphi = \xi$ has a unique solution for every $\xi \in L^2(\mathcal{M})$ **or** there exists a non-zero solution of $\mathcal{T}\varphi_0 = 0$, with an eigenvector of $\mathcal{K}$ corresponding to the eigenvalue 1. The theory remains the same if instead of $\mathcal{T}$ we consider the operator $\mathcal{T}_\lambda = \mathbb{1} - \lambda\mathcal{K}$ with $\lambda \neq 0$. As $\mathcal{K}$ is a compact operator there is at most a denumerable set of λ for which the second part of the Fredholm alternative holds: apart from this set the inverse operator $(\mathbb{1} - \lambda\mathcal{T})^{-1}$ exists and is bounded (in the operator sense). When λ is sufficiently small we may look for a perturbative expression for such an inverse, as a geometric series

$$(\mathbb{1} - \lambda\mathcal{K})^{-1} = \mathbb{1} + \lambda\mathcal{K} + \lambda^2\mathcal{K}^2 + \dots = \mathbb{1} + \lambda\mathcal{W}, \quad (28.7)$$

remark A22.1

where $\mathcal{K}^n$ is a compact integral operator with kernel

$$\mathcal{K}^n(x, y) = \int_{\mathcal{M}^{n-1}} dz_1 \dots dz_{n-1} \mathcal{K}(x, z_1) \dots \mathcal{K}(z_{n-1}, y),$$

and $\mathcal{W}$ is also compact, as it is given by the convergent sum of compact operators. The problem with (28.7) is that the series has a finite radius of convergence, while apart from a denumerable set of λ 's the inverse operator is well defined. A fundamental result in the theory of integral equations consists in rewriting the resolving kernel $\mathcal{W}$ as a ratio of two *analytic* functions of λ

$$\mathcal{W}(x, y) = \frac{\mathcal{D}(x, y; \lambda)}{D(\lambda)}.$$

If we introduce the notation

$$\mathcal{K} \begin{pmatrix} x_1 \dots x_n \\ y_1 \dots y_n \end{pmatrix} = \begin{vmatrix} \mathcal{K}(x_1, y_1) & \dots & \mathcal{K}(x_1, y_n) \\ \dots & \dots & \dots \\ \mathcal{K}(x_n, y_1) & \dots & \mathcal{K}(x_n, y_n) \end{vmatrix}$$

we may write the explicit expressions

$$\begin{aligned} D(\lambda) &= 1 + \sum_{n=1}^{\infty} (-1)^n \frac{\lambda^n}{n!} \int_{\mathcal{M}^n} dz_1 \dots dz_n \mathcal{K} \begin{pmatrix} z_1 \dots z_n \\ z_1 \dots z_n \end{pmatrix} \\ &= \exp \left(- \sum_{m=1}^{\infty} \frac{\lambda^m}{m} \text{Tr} \mathcal{K}^m \right) \\ \mathcal{D}(x, y; \lambda) &= \mathcal{K} \begin{pmatrix} x \\ y \end{pmatrix} + \sum_{n=1}^{\infty} \frac{(-\lambda)^n}{n!} \int_{\mathcal{M}^n} dz_1 \dots dz_n \mathcal{K} \begin{pmatrix} x & z_1 & \dots & z_n \\ y & z_1 & \dots & z_n \end{pmatrix} \end{aligned} \quad (28.8)$$

The quantity $D(\lambda)$ is known as the Fredholm determinant (see (22.16)): it is an entire analytic function of λ , and $D(\lambda) = 0$ if and only if $1/\lambda$ is an eigenvalue of $\mathcal{K}$.

Worth emphasizing again: the Fredholm theory is based on the compactness of the integral operator, i.e., on the functional properties (summability) of its kernel. As the Perron-Frobenius operator is not compact, there is a bit of wishful thinking involved here.

28.4 Analyticity of spectral determinants

They savored the strange warm glow of being much more ignorant than ordinary people, who were only ignorant of ordinary things.

—Terry Pratchett

Spaces of functions integrable L^1 , or square-integrable L^2 on interval $[0, 1]$ are mapped into themselves by the Perron-Frobenius operator, and in both cases the constant function $\phi_0 \equiv 1$ is an eigenfunction with eigenvalue 1. If we focus our attention on L^1 we also have a family of L^1 eigenfunctions,

$$\phi_\theta(y) = \sum_{k \neq 0} \exp(2\pi i k y) \frac{1}{|k|^\theta} \quad (28.9)$$

with complex eigenvalue $2^{-\theta}$, parameterized by complex θ with $\text{Re } \theta > 0$. By varying θ one realizes that such eigenvalues fill out the entire unit disk. Such *essential spectrum*, the case $k = 0$ of figure 28.5, hides all fine details of the spectrum.

What's going on? Spaces L^1 and L^2 contain arbitrarily ugly functions, allowing any singularity as long as it is (square) integrable - and there is no way that expanding dynamics can smooth a kinky function with a non-differentiable singularity, let's say a discontinuous step, and that is why the eigenspectrum is dense rather than discrete. Mathematicians love to wallow in this kind of muck, but there is no way to prepare a nowhere differentiable L^1 initial density in a laboratory. The only thing we can prepare and measure are piecewise smooth (real-analytic) density functions.

For a bounded linear operator $\mathcal{A}$ on a Banach space Ω , the spectral radius is the smallest positive number ρ_{spec} such that the spectrum is inside the disk of radius ρ_{spec} , while the essential spectral radius is the smallest positive number ρ_{ess} such that outside the disk of radius ρ_{ess} the spectrum consists only of isolated eigenvalues of finite multiplicity (see figure 28.5).

exercise 28.5

We may shrink the essential spectrum by letting the Perron-Frobenius operator act on a space of smoother functions, exactly as in the one-branch repeller case of sect. 28.1. We thus consider a smaller space, $\mathbb{C}^{k+\alpha}$, the space of k times

differentiable functions whose k 'th derivatives are Hölder continuous with an exponent $0 < \alpha \leq 1$: the expansion property guarantees that such a space is mapped into itself by the Perron-Frobenius operator. In the strip $0 < \operatorname{Re} \theta < k + \alpha$ most ϕ_θ will cease to be eigenfunctions in the space $\mathbb{C}^{k+\alpha}$; the function ϕ_n survives only for integer valued $\theta = n$. In this way we arrive at a finite set of *isolated* eigenvalues $1, 2^{-1}, \dots, 2^{-k}$, and an essential spectral radius $\rho_{ess} = 2^{-(k+\alpha)}$.

We follow a simpler path and restrict the function space even further, namely to a space of analytic functions, i.e., functions for which the Taylor expansion is convergent at each point of the interval $[0, 1]$. With this choice things turn out easy and elegant. To be more specific, let ϕ be a holomorphic and bounded function on the disk $D = B(0, R)$ of radius $R > 0$ centered at the origin. Our Perron-Frobenius operator preserves the space of such functions provided $(1 + R)/2 < R$ so all we need is to choose $R > 1$. If F_s , $s \in \{0, 1\}$, denotes the s inverse branch of the Bernoulli shift (28.19), the corresponding action of the Perron-Frobenius operator is given by $\mathcal{L}_s h(y) = \sigma F'_s(y) h \circ F_s(y)$, using the Cauchy integral formula along the ∂D boundary contour:

$$\mathcal{L}_s h(y) = \sigma \oint_{\partial D} \frac{dw}{2\pi i} \frac{h(w)F'_s(y)}{w - F_s(y)}. \quad (28.10)$$

For reasons that will be made clear later we have introduced a sign $\sigma = \pm 1$ of the given real branch $|F'(y)| = \sigma F'(y)$. For both branches of the Bernoulli shift $\sigma = 1$, but in general one is not allowed to take absolute values as this could destroy analyticity. In the above formula one may also replace the domain D by *any domain* containing $[0, 1]$ such that the inverse branches maps the closure of D into the interior of D . Why? simply because the kernel remains non-singular under this condition, i.e., $w - F(y) \neq 0$ whenever $w \in \partial D$ and $y \in \operatorname{Cl} D$. The problem is now reduced to the standard theory for Fredholm determinants, sect. 28.3. The integral kernel is no longer singular, traces and determinants are well-defined, and we can evaluate the trace of $\mathcal{L}_F$ by means of the Cauchy contour integral formula:

$$\operatorname{tr} \mathcal{L}_F = \oint \frac{dw}{2\pi i} \frac{\sigma F'(w)}{w - F(w)}.$$

Elementary complex analysis shows that since F maps the closure of D into its own interior, F has a unique (real-valued) fixed point x^* with a multiplier strictly smaller than one in absolute value. Residue calculus therefore yields

$$\operatorname{tr} \mathcal{L}_F = \frac{\sigma F'(x^*)}{1 - F'(x^*)} = \frac{1}{|f'(x^*) - 1|},$$

justifying our previous *ad hoc* calculations of traces using Dirac delta functions.

We worked out a very specific example, yet our conclusions can be generalized, provided a number of restrictive requirements are met by the dynamical system under investigation:

- 1) the evolution operator is *multiplicative* along the flow,
- 2) the symbolic dynamics is a *finite subshift*,

exercise 28.6

exercise 28.6

- 3) all cycle eigenvalues are *hyperbolic* (exponentially bounded in magnitude away from 1),
- 4) the map (or the flow) is *real analytic*, i.e., it has a piecewise analytic continuation to a complex extension of the state space.

These assumptions are romantic expectations not satisfied by the dynamical systems that we actually desire to understand. Still, they are not devoid of physical interest; for example, nice repellers like our 3-disk game of pinball do satisfy the above requirements.

Properties 1 and 2 enable us to represent the evolution operator as a finite matrix in an appropriate basis; properties 3 and 4 enable us to bound the size of the matrix elements and control the eigenvalues. To see what can go wrong, consider the following examples:

Property 1 is violated for flows in 3 or more dimensions by the following weighted evolution operator

$$\mathcal{L}^t(y, x) = |\Lambda^t(x)|^\beta \delta(y - f^t(x)) ,$$

where $\Lambda^t(x)$ is an eigenvalue of the Jacobian matrix transverse to the flow. Semi-classical quantum mechanics suggest operators of this form with $\beta = 1/2$. The problem with such operators arises from the fact that when considering the Jacobian matrices $J_{ab} = J_a J_b$ for two successive trajectory segments a and b , the corresponding eigenvalues are in general *not* multiplicative, $\Lambda_{ab} \neq \Lambda_a \Lambda_b$ (unless a, b are iterates of the same prime cycle p , so $J_a J_b = J_p^{r_a+r_b}$). Consequently, this evolution operator is not multiplicative along the trajectory. The theorems require that the evolution be represented as a matrix in an appropriate polynomial basis, and thus cannot be applied to non-multiplicative kernels, i.e., kernels that do not satisfy the semi-group property $\mathcal{L}^t \mathcal{L}^t = \mathcal{L}^{t'+t}$.

Property 2 is violated by the 1-dimensional tent map (see figure 28.3 (a))

$$f(x) = \alpha(1 - |1 - 2x|) , \quad 1/2 < \alpha < 1 .$$

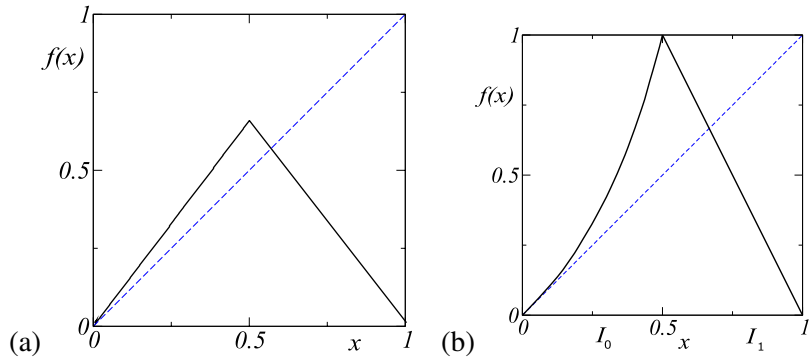
All cycle eigenvalues are hyperbolic, but in general the critical point $x_c = 1/2$ is not a pre-periodic point, so there is no finite Markov partition and the symbolic dynamics does not have a finite grammar (see sect. 15.4 for definitions). In practice, this means that while the leading eigenvalue of $\mathcal{L}$ might be computable, the rest of the spectrum is very hard to control; as the parameter α is varied, the non-leading zeros of the spectral determinant move wildly about.

Property 3 is violated by the map (see figure 28.3 (b))

$$f(x) = \begin{cases} x + 2x^2 & , \quad x \in I_0 = [0, \frac{1}{2}] \\ 2 - 2x & , \quad x \in I_1 = [\frac{1}{2}, 1] \end{cases} .$$

Here the interval $[0, 1]$ has a Markov partition into two subintervals I_0 and I_1 , and f is monotone on each. However, the fixed point at $x = 0$ has marginal stability $\Lambda_0 = 1$, and violates condition 3. This type of map is called “intermittent” and

Figure 28.3: (a) A (hyperbolic) tent map without a finite Markov partition. (b) A Markov map with a marginal fixed point.



necessitates much extra work. The problem is that the dynamics in the neighborhood of a marginal fixed point is very slow, with correlations decaying as power laws rather than exponentially. We will discuss such flows in chapter 29.

Property 4 is required as the heuristic approach of chapter 21 faces two major hurdles:

1. The trace (21.7) is not well defined because the integral kernel is singular.
2. The existence and properties of eigenvalues are by no means clear.

Actually, property 4 is quite restrictive, but we need it in the present approach, so that the Banach space of analytic functions in a disk is preserved by the Perron-Frobenius operator.

In attempting to generalize the results, we encounter several problems. First, in higher dimensions life is not as simple. Multi-dimensional residue calculus is at our disposal but in general requires that we find poly-domains (direct product of domains in each coordinate) and this need not be the case. Second, and perhaps somewhat surprisingly, the ‘counting of periodic orbits’ presents a difficult problem. For example, instead of the Bernoulli shift consider the doubling map (14.19) of the circle, $x \mapsto 2x \bmod 1$, $x \in \mathbb{R}/\mathbb{Z}$. Compared to the shift on the interval $[0, 1]$ the only difference is that the endpoints 0 and 1 are now glued together. Because these endpoints are fixed points of the map, the number of cycles of period n decreases by 1. The determinant becomes:

$$\det(1 - z\mathcal{L}) = \exp\left(-\sum_{n=1}^{\infty} \frac{z^n}{n} \frac{2^n - 1}{2^n - 1}\right) = 1 - z. \quad (28.11)$$

The value $z = 1$ still comes from the constant eigenfunction, but the Bernoulli polynomials no longer contribute to the spectrum (as they are not periodic). Proofs of these facts, however, are difficult if one sticks to the space of analytic functions.

Third, our Cauchy formulas *a priori* work only when considering purely expanding maps. When stable and unstable directions co-exist we have to resort to stranger function spaces, as shown in the next section.

28.5 Hyperbolic maps

I can give you a definition of a Banach space, but I do not know what that means.

—Federico Bonnetto, *Banach space*

(H.H. Rugh)

Proceeding to hyperbolic systems, one faces the following paradox: If f is an area-preserving hyperbolic and real-analytic map of, for example, a 2-dimensional torus then the Perron-Frobenius operator is unitary on the space of L^2 functions, and its spectrum is confined to the unit circle. On the other hand, when we compute determinants we find eigenvalues scattered around inside the unit disk. Thinking back to the Bernoulli shift example 28.2, one would like to imagine these eigenvalues as popping up from the L^2 spectrum by shrinking the function space. Shrinking the space, however, can only make the spectrum smaller so this is obviously not what happens. Instead one needs to introduce a ‘mixed’ function space where in the unstable direction one resorts to analytic functions, as before, but in the stable direction one instead considers a ‘dual space’ of distributions on analytic functions. Such a space is neither included in nor includes L^2 and we have thus resolved the paradox. However, it still remains to be seen how traces and determinants are calculated.

The linear hyperbolic fixed point example 28.6 is somewhat misleading, as we have made explicit use of a map that acts independently along the stable and unstable directions. For a more general hyperbolic map, there is no way to implement such direct product structure, and the whole argument falls apart. Here comes an idea; use the analyticity of the map to rewrite the Perron-Frobenius operator acting as follows (where σ denotes the sign of the derivative in the unstable direction):

$$\mathcal{L}h(z_1, z_2) = \oint \oint \frac{\sigma h(w_1, w_2)}{(z_1 - f_1(w_1, w_2)(f_2(w_1, w_2) - z_2))} \frac{dw_1}{2\pi i} \frac{dw_2}{2\pi i}. \quad (28.12)$$

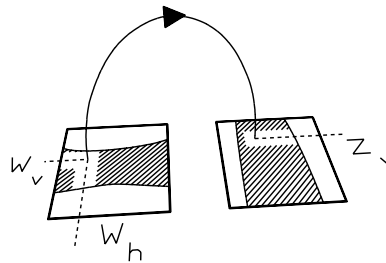
Here the function ϕ should belong to a space of functions analytic respectively *outside* a disk and *inside* a disk in the first and the second coordinates; with the additional property that the function decays to zero as the first coordinate tends to infinity. The contour integrals are along the boundaries of these disks. It is an exercise in multi-dimensional residue calculus to verify that for the above linear example this expression reduces to (28.25). Such operators form the building blocks in the calculation of traces and determinants. One can prove the following:

Theorem: *The spectral determinant for 2-dimensional hyperbolic analytic maps is entire.*

remark 28.8

The proof, apart from the Markov property that is the same as for the purely expanding case, relies heavily on the analyticity of the map in the explicit construction of the function space. The idea is to view the hyperbolicity as a cross product of a contracting map in forward time and another contracting map in backward time. In this case the Markov property introduced above has to be elaborated

Figure 28.4: An analytic hyperbolic map that defines a unique trajectory between two rectangles, where w_h is the contracting coordinate at the initial rectangle and z_v is the expanding coordinate at the image rectangle. In particular, w_v and z_h (not shown) are uniquely specified.



a bit. Instead of dividing the state space into intervals, one divides it into rectangles. The rectangles should be viewed as a direct product of intervals (say horizontal and vertical), such that the forward map is contracting in, for example, the horizontal direction, while the inverse map is contracting in the vertical direction. For Axiom A systems (see remark 28.8) one may choose coordinate axes close to the stable/unstable manifolds of the map. With the state space divided into N rectangles $\{\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_N\}$, $\mathcal{M}_i = I_i^h \times I_i^v$ one needs a complex extension $D_i^h \times D_i^v$, with which the hyperbolicity condition (which simultaneously guarantees the Markov property) can be formulated as follows:

Analytic hyperbolic property: Either $f(\mathcal{M}_i) \cap \text{Int}(\mathcal{M}_j) = \emptyset$, or for each pair $w_h \in \text{Cl}(D_i^h)$, $z_v \in \text{Cl}(D_j^v)$ there exist unique analytic functions of w_h, z_v : $w_v = w_v(w_h, z_v) \in \text{Int}(D_i^v)$, $z_h = z_h(w_h, z_v) \in \text{Int}(D_j^h)$, such that $f(w_h, w_v) = (z_h, z_v)$. Furthermore, if $w_h \in I_i^h$ and $z_v \in I_j^v$, then $w_v \in I_i^v$ and $z_h \in I_j^h$ (see figure 28.4).

In plain English, this means for the iterated map that one replaces the coordinates z_h, z_v at time n by the contracting pair z_h, w_v , where w_v is the contracting coordinate at time $n + 1$ for the ‘partial’ inverse map.

In two dimensions the operator in (28.12) acts on functions analytic outside D_i^h in the horizontal direction (and tending to zero at infinity) and inside D_i^v in the vertical direction. The contour integrals are precisely along the boundaries of these domains.

A map f satisfying the above condition is called *analytic hyperbolic* and the theorem states that the associated spectral determinant is entire, and that the trace formula (21.7) is correct.

Examples of analytic hyperbolic maps are provided by small analytic perturbations of the cat map, the 3-disk repeller, and the 2-dimensional baker’s map.

28.6 Physics of eigenvalues and eigenfunctions



By now we appreciate that any honest attempt to look at the spectral properties of the Perron-Frobenius operator involves hard mathematics, but the reward is of this effort is that we are able to control the analyticity properties of dynamical

zeta functions and spectral determinants, and thus substantiate the claim that these objects provide a powerful and well-founded theory.

Often (see chapter 20) physically important part of the spectrum is just the leading eigenvalue, which gives us the escape rate from a repeller, or, for a general evolution operator, formulas for expectation values of observables and their higher moments. Also the eigenfunction associated to the leading eigenvalue has a physical interpretation (see chapter 19): it is the density of the natural measures, with singular measures ruled out by the proper choice of the function space. This conclusion is in accord with the generalized Perron-Frobenius theorem for evolution operators. In a finite dimensional setting, the statement is:

remark 28.7

- **Perron-Frobenius theorem:** Let L_{ij} be a non-negative matrix, such that some finite n exists for which any initial state has reached any other state, $(L^n)_{ij} > 0 \forall i, j$: then

1. The maximal modulus eigenvalue is non-degenerate, real, and positive,
2. The corresponding eigenvector (defined up to a constant) has non-negative coordinates.

We may ask what physical information is contained in eigenvalues beyond the leading one: suppose that we have a probability conserving system (so that the dominant eigenvalue is 1), for which the essential spectral radius satisfies $0 < \rho_{ess} < \theta < 1$ on some Banach space $\mathcal{B}$. Denote by $\mathbf{P}$ the projection corresponding to the part of the spectrum inside a disk of radius θ . We denote by $\lambda_1, \lambda_2, \dots, \lambda_M$ the eigenvalues outside of this disk, ordered by the size of their absolute value, with $\lambda_1 = 1$. Then we have the following decomposition

$$\mathcal{L}\varphi = \sum_{i=1}^M \lambda_i \psi_i L_i \psi_i^* \varphi + \mathbf{P}\mathcal{L}\varphi \quad (28.13)$$

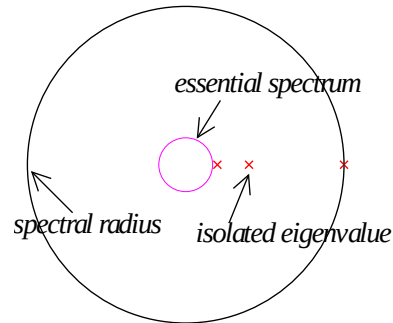
when L_i are (finite) matrices in Jordan canonical form ($L_0 = 0$ is a $[1 \times 1]$ matrix, as λ_0 is simple, due to the Perron-Frobenius theorem), whereas ψ_i is a row vector whose elements form a basis on the eigenspace corresponding to λ_i , and ψ_i^* is a column vector of elements of $\mathcal{B}^*$ (the dual space of linear functionals over $\mathcal{B}$) spanning the eigenspace of $\mathcal{L}^*$ corresponding to λ_i . For iterates of the Perron-Frobenius operator, (28.13) becomes

$$\mathcal{L}^n \varphi = \sum_{i=1}^M \lambda_i^n \psi_i L_i^n \psi_i^* \varphi + \mathbf{P}\mathcal{L}^n \varphi. \quad (28.14)$$

If we now consider, for example, correlation between initial φ evolved n steps and final ξ ,

$$\langle \xi | \mathcal{L}^n | \varphi \rangle = \int_{\mathcal{M}} dy \xi(y) (\mathcal{L}^n \varphi)(y) = \int_{\mathcal{M}} dw (\xi \circ f^n)(w) \varphi(w), \quad (28.15)$$

Figure 28.5: Spectrum of the Perron-Frobenius operator acting on the space of $\mathbb{C}^{k+\alpha}$ Hölder-continuous functions: only k isolated eigenvalues remain between the spectral radius, and the essential spectral radius which bounds the “essential,” continuous spectrum.



it follows that

$$\langle \xi | \mathcal{L}^n | \varphi \rangle = \lambda_1^n \omega_1(\xi, \varphi) + \sum_{i=2}^L \lambda_i^n \omega_i^{(n)}(\xi, \varphi) + O(\theta^n), \quad (28.16)$$

where

$$\omega_i^{(n)}(\xi, \varphi) = \int_{\mathcal{M}} dy \xi(y) \psi_i L_i^n \psi_i^* \varphi.$$

The eigenvalues beyond the leading one provide two pieces of information: they rule the convergence of expressions containing high powers of the evolution operator to leading order (the λ_1 contribution). Moreover if $\omega_1(\xi, \varphi) = 0$ then (28.15) defines a correlation function: as each term in (28.16) vanishes exponentially in the $n \rightarrow \infty$ limit, the eigenvalues $\lambda_2, \dots, \lambda_M$ determine the exponential decay of correlations for our dynamical system. The prefactors ω depend on the choice of functions, whereas the exponential decay rates (given by logarithms of λ_i) do not: the correlation spectrum is thus a *universal* property of the dynamics (once we fix the overall functional space on which the Perron-Frobenius operator acts).

exercise 28.7

 example 28.8
p. 578

28.7 Troubles ahead

... just a bunch of flawed attempts by imperfect humans to
make sense of a world besieged by chaos.

— Daniela Witten

The above discussion confirms that for a series of examples of increasing generality formal manipulations with traces and determinants are justified: the Perron-Frobenius operator has isolated eigenvalues, the trace formulas are explicitly verified, and the spectral determinant is an entire function whose zeroes yield the eigenvalues. Real life is harder, as we may appreciate through the following considerations:

- Our discussion tacitly assumed something that is physically entirely reasonable: our evolution operator is acting on the space of analytic functions, i.e., we are allowed to represent the initial density $\rho(x)$ by its Taylor expansions in the neighborhoods of periodic points. This is however far from being the only possible choice: mathematicians often work with the function space $\mathbb{C}^{k+\alpha}$, i.e., the space of k times differentiable functions whose k 'th derivatives are Hölder continuous with an exponent $0 < \alpha \leq 1$: then every y^η with $\operatorname{Re} \eta > k$ is an eigenfunction of the Perron-Frobenius operator and we have

exercise 28.1

$$\mathcal{L}y^\eta = \frac{1}{|\Lambda|\Lambda^\eta}y^\eta, \quad \eta \in \mathbb{C}.$$

This spectrum differs markedly from the analytic case: only a small number of isolated eigenvalues remain, enclosed between the spectral radius and a smaller disk of radius $1/|\Lambda|^{k+1}$, see figure 28.5. In literature the radius of this disk is called the *essential spectral radius*.

In sect. 28.4 we discussed this point further, with the aid of a less trivial 1-dimensional example. The physical point of view is complementary to the standard setting of ergodic theory, where many chaotic properties of a dynamical system are encoded by the presence of a *continuous* spectrum, used to prove asymptotic decay of correlations in the space of L^2 square-integrable functions.

exercise 28.2

- A deceptively innocent assumption is hidden beneath much that was discussed so far: that (28.17) maps a given function space into itself. The *expanding* property of the map guarantees that: if $f(x)$ is smooth in a domain D then $f(x/\Lambda)$ is smooth on a *larger* domain, provided $|\Lambda| > 1$. For higher-dimensional hyperbolic flows this is not the case, and, as we saw in sect. 28.5, extensions of the results obtained for expanding 1-dimensional maps are highly nontrivial.
- It is not at all clear that the above analysis of a simple one-branch, one fixed point repeller can be extended to dynamical systems with Cantor sets of periodic points: we showed this in sect. 28.4.

Résumé

Examples of analytic eigenfunctions for 1-dimensional maps are seductive, and make the problem of evaluating ergodic averages appear easy; just integrate over the desired observable weighted by the natural measure, right? No, generic natural measure sits on a fractal set and is singular everywhere. The point of ChaosBook is that you never need to construct the natural measure, cycle expansions will do that job.

A theory of evaluation of dynamical averages by means of trace formulas and spectral determinants requires a deep understanding of their analyticity and convergence. We worked here through a series of examples:

1. exact spectrum (but for a single fixed point of a linear map)
2. exact spectrum for a locally analytic map, matrix representation
3. rigorous proof of existence of discrete spectrum for 2-dimensional hyperbolic maps

In the case of especially well-behaved “Axiom A” systems, where both the symbolic dynamics and hyperbolicity are under control, it is possible to treat traces and determinants in a rigorous fashion, and strong results about the analyticity properties of dynamical zeta functions and spectral determinants outlined above follow.

Most systems of interest are *not* of the “axiom A” category; they are neither purely hyperbolic nor (as we have seen in chapters 14 and 15) do they have finite grammar. The importance of symbolic dynamics is generally grossly under appreciated; the crucial ingredient for nice analyticity properties of zeta functions is the existence of a finite grammar (coupled with uniform hyperbolicity).

The dynamical systems which are *really* interesting - for example, smooth bounded Hamiltonian potentials - are presumably never fully chaotic, and the central question remains: How do we attack this problem in a systematic and controllable fashion?

Theorem: Conjecture 3 with technical hypothesis is true in a lot of cases.

— M. Shub

Commentary

Remark 28.1. Surveys of rigorous theory. We recommend the references listed in remark 1.1 for an introduction to the mathematical literature on this subject. For a physicist, Driebe’s monograph [8] might be the most accessible introduction into mathematics discussed briefly in this chapter. There are a number of reviews of the mathematical approach to dynamical zeta functions and spectral determinants, with pointers to the original references, such as refs. [3, 17]. An alternative approach to spectral properties of the Perron-Frobenius operator is given in ref. [24].

Ergodic theory, as presented by Sinai [22] and others, tempts one to describe the densities on which the evolution operator acts in terms of either integrable or square-integrable functions. For our purposes, as we have already seen, this space is not suitable. An introduction to ergodic theory is given by Sinai, Kornfeld and Fomin [15]; more advanced old-fashioned presentations are Walters [25] and Denker, Grillenberger and Sigmond [6]; and a more formal one is given by Peterson [16].

Remark 28.2. Fredholm theory. Our brief summary of Fredholm theory is based on the exposition of ref. [14]. A technical introduction of the theory from an operator point of view is given in ref. [7]. The theory is presented in a more general form in ref. [12].

Remark 28.3. Bernoulli shift. For a more in-depth discussion, consult chapter 3 of ref. [8]. The extension of Fredholm theory to the case of Bernoulli shift on $\mathbb{C}^{k+\alpha}$ (in which the Perron-Frobenius operator is *not* compact – technically it is only *quasi-compact*. That is, the essential spectral radius is strictly smaller than the spectral radius) has been given by Ruelle [20]: a concise and readable statement of the results is contained in ref. [2]. We see from (28.29) that for the Bernoulli shift the exponential decay rate of correlations coincides with the Lyapunov exponent: while such an identity holds for a number of systems, it is by no means a general result, and there exist explicit counterexamples. See also remark 14.2.

Remark 28.4. Hyperbolic dynamics. When dealing with hyperbolic systems one might try to reduce to the expanding case by projecting the dynamics along the unstable directions. As mentioned in the text this can be quite involved technically, as such unstable foliations are not characterized by strong smoothness properties. For such an approach, see ref. [24].

Remark 28.5. Spectral determinants for smooth flows. The theorem on page 566 also applies to hyperbolic analytic maps in d dimensions and smooth hyperbolic analytic flows in $(d + 1)$ dimensions, provided that the flow can be reduced to a piecewise analytic map by a suspension on a Poincaré section, complemented by an analytic “ceiling” function (3.7) that accounts for a variation in the section return times. For example, if we take as the ceiling function $g(x) = e^{sT(x)}$, where $T(x)$ is the next Poincaré section time for a trajectory starting at x , we reproduce the flow spectral determinant (22.23). Proofs are beyond the scope of this chapter.

Remark 28.6. Explicit diagonalization. For 1-dimensional repellers a diagonalization of an explicit truncated $\mathbf{L}_{mn}$ matrix evaluated in a judiciously chosen basis may yield many more eigenvalues than a cycle expansion (see refs. [1, 5]). The reasons why one persists in using periodic orbit theory are partially aesthetic and partially pragmatic. The explicit calculation of $\mathbf{L}_{mn}$ demands an explicit choice of a basis and is thus non-invariant, in contrast to cycle expansions which utilize only the invariant information of the flow. In addition, we usually do not know how to construct $\mathbf{L}_{mn}$ for a realistic high-dimensional flow, such as the hyperbolic 3-disk game of pinball flow of sect. 1.3, whereas periodic orbit theory is true in higher dimensions and straightforward to apply.

Remark 28.7. Perron-Frobenius theorem. A proof of the Perron-Frobenius theorem may be found in ref. [25]. For positive transfer operators, this theorem has been generalized by Ruelle [18].

Remark 28.8. Axiom A systems. The proofs in sect. 28.5 follow the thesis work of H.H. Rugh [9, 19, 21]. For a mathematical introduction to the subject, consult the excellent review by V. Baladi [3]. It would take us too far afield to give and explain the definition of Axiom A systems (see refs. [4, 23]). Axiom A implies, however, the existence of a Markov partition of the state space from which the properties 2 and 3 assumed on page 575 follow.

Remark 28.9. Left eigenfunctions. We shall never use an explicit form of left eigenfunctions, corresponding to highly singular kernels like (28.31). Many details have been elaborated in a number of papers, such as ref. [13], with a daring physical interpretation.

Remark 28.10. Ulam's idea. The approximation of Perron-Frobenius operator defined by (19.12) has been shown to reproduce the spectrum for expanding maps, once finer and finer Markov partitions are used [10]. The subtle point of choosing a state space partitioning for a “generic case” is discussed in ref. [11].

References

- [1] D. Alonso, P. G. D. MacKernan, and G. Nicolis, “Statistical approach to nonhyperbolic chaotic systems”, *Phys. Rev. E* **54**, 2474 (1996).
- [2] V. Baladi, Dynamical zeta functions, in *Real and Complex Dynamical Systems: Proceedings of the NATO ASI*, edited by B. Branner and P. Hjorth (1995), pp. 1–26.
- [3] V. Baladi, “A brief introduction to dynamical zeta functions”, in *Classical Nonintegrability, Quantum Chaos*, edited by A. Knauf and Y. G. Sinai (Springer, 1997), pp. 3–20.
- [4] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms* (Springer, New York, 1975).
- [5] F. Christiansen, P. Cvitanović, and H. H. Rugh, “The spectrum of the period-doubling operator in terms of cycles”, *J. Phys. A* **23**, L713S–L717S (1990).
- [6] M. Denker, C. Grillenberger, and K. Sigmund, *Ergodic Theory on Compact Spaces* (Springer, 1976).

- [7] G. Douglas, *Banach Algebra Techniques in Operator Theory* (Springer, New York, 1998).
- [8] D. J. Driebe, *Fully Chaotic Maps and Broken Time Symmetry* (Springer, New York, 1999).
- [9] D. Fried, “The Zeta functions of Ruelle and Selberg I”, *Ann. Scient. Ec. Norm. Sup.* **19**, 491 (1986).
- [10] G. Froyland, “Computer-assisted bounds for the rate of decay of correlations”, *Commun. Math. Phys.* **189**, 237–257 (1997).
- [11] G. Froyland, Extracting dynamical behaviour via Markov models, in *Non-linear Dynamics and Statistics: Proc. Newton Institute, Cambridge 1998*, edited by A. Mees (2001), pp. 281–321.
- [12] A. Grothendieck, “La théorie de Fredholm”, *Bull. Soc. Math. France* **84**, 319–384 (1956).
- [13] H. H. Hasegawa and W. C. Saphir, “Unitarity and irreversibility in chaotic systems”, *Phys. Rev. A* **46**, 7401–7423 (1992).
- [14] A. N. Kolmogorov and S. V. Fomin, *Elements of the Theory of Functions and Functional Analysis* (Dover, New York, 1980).
- [15] I. Kornfeld, S. Fomin, and Y. G. Sinai, *Ergodic Theory* (Springer, New York, 1982).
- [16] K. Peterson, *Ergodic Theory* (Cambridge Univ. Press, Cambridge UK, 1990).
- [17] M. Pollicott, “Periodic orbits and zeta functions”, in *Handbook of Dynamical Systems*, Vol. 1, Part A, edited by B. Hasselblatt and A. Katok (Elsevier, New York, 2002), pp. 409–452.
- [18] D. Ruelle, “Statistical mechanics of a one-dimensional lattice gas”, *Commun. Math. Phys.* **9**, 267–288 (1968).
- [19] D. Ruelle, “Zeta-functions for expanding maps and Anosov flows”, *Inv. Math.* **34**, 231–242 (1976).
- [20] D. Ruelle, “An extension of the theory of Fredholm determinants”, *Inst. Hautes Études Sci. Publ. Math.* **72**, 175–193 (1990).
- [21] H. H. Rugh, “The correlation spectrum for hyperbolic analytic maps”, *Nonlinearity* **5**, 1237 (1992).
- [22] Y. G. Sinai, *Topics in Ergodic Theory* (Princeton Univ. Press, Princeton, 1994).
- [23] S. Smale, “Differentiable dynamical systems”, *Bull. Amer. Math. Soc.* **73**, 747–817 (1967).
- [24] M. Viana, *Stochastic Dynamics of Deterministic Systems*, Vol. 21 (IMPA, 1997).
- [25] P. Walters, *An Introduction to Ergodic Theory* (Springer, New York, 1981).
- [26] A. Zygmund, *Trigonometric Series* (Cambridge Univ. Press, Cambridge UK, 1959).

28.8 Examples

Example 28.1. The simplest eigenspectrum - a single fixed point: In order to get some feeling for the determinants defined so formally in sect. 22.2, let us work out a trivial example: a repeller with only one expanding linear branch

$$f(x) = \Lambda x, \quad |\Lambda| > 1,$$

and only one fixed point $x_q = 0$. The action of the Perron-Frobenius operator (19.10) is

$$\mathcal{L}\phi(y) = \int dx \delta(y - \Lambda x) \phi(x) = \frac{1}{|\Lambda|} \phi(y/\Lambda). \quad (28.17)$$

From this one immediately gets that the monomials y^k are eigenfunctions:

[return to p. 556](#)

$$\mathcal{L}y^k = \frac{1}{|\Lambda|\Lambda^k} y^k, \quad k = 0, 1, 2, \dots \quad (28.18)$$

Example 28.2. Eigenfunction of Bernoulli shift map. (continued from example 14.5) The Bernoulli shift map figure 28.1

$$f(x) = \begin{cases} f_0(x) = 2x, & x \in I_0 = [0, 1/2) \\ f_1(x) = 2x - 1, & x \in I_1 = (1/2, 1] \end{cases} \quad (28.19)$$

models the 50-50% probability of a coin toss. The associated Perron-Frobenius operator (19.9) assembles $\rho(y)$ from its two preimages

$$\mathcal{L}\rho(y) = \frac{1}{2}\rho\left(\frac{y}{2}\right) + \frac{1}{2}\rho\left(\frac{y+1}{2}\right). \quad (28.20)$$

For this simple example the eigenfunctions can be written down explicitly: they coincide, up to constant prefactors, with the Bernoulli polynomials $B_n(x)$. These polynomials are generated by the Taylor expansion of the exponential generating function

$$\mathcal{G}(x, t) = \frac{te^{xt}}{e^t - 1} = \sum_{k=0}^{\infty} B_k(x) \frac{t^k}{k!}, \quad B_0(x) = 1, \quad B_1(x) = x - \frac{1}{2}, \dots$$

The Perron-Frobenius operator (28.20) acts on the exponential generating function $\mathcal{G}$ as

$$\mathcal{L}\mathcal{G}(x, t) = \frac{1}{2} \left(\frac{te^{xt/2}}{e^t - 1} + \frac{te^{t/2}e^{xt/2}}{e^t - 1} \right) = \frac{t}{2} \frac{e^{xt/2}}{e^{t/2} - 1} = \sum_{k=1}^{\infty} B_k(x) \frac{(t/2)^k}{k!},$$

hence each $B_k(x)$ is an eigenfunction of $\mathcal{L}$ with eigenvalue $1/2^k$.

The full operator has two components corresponding to the two branches. For the n times iterated operator we have a full binary shift, and for each of the 2^n branches the above calculations carry over, yielding the same trace $(2^n - 1)^{-1}$ for every cycle on period n . Without further ado we substitute everything back and obtain the determinant,

[return to p. 557](#)

$$\det(1 - z\mathcal{L}) = \exp\left(-\sum_{n=1}^{\infty} \frac{z^n}{n} \frac{2^n}{2^n - 1}\right) = \prod_{k=0}^{\infty} \left(1 - \frac{z}{2^k}\right), \quad (28.21)$$

verifying that the Bernoulli polynomials are eigenfunctions with eigenvalues $1, 1/2, \dots, 1/2^n, \dots$.

Example 28.3. The trace formula for a single fixed point: The eigenvalues Λ^{-k-1} fall off exponentially with k , so the trace of $\mathcal{L}$ is a convergent sum

$$\mathrm{tr} \mathcal{L} = \frac{1}{|\Lambda|} \sum_{k=0}^{\infty} \Lambda^{-k} = \frac{1}{|\Lambda|(1 - \Lambda^{-1})} = \frac{1}{|f(0)' - 1|},$$

in agreement with (21.6). A similar result follows for powers of $\mathcal{L}$, yielding the single-fixed point version of the trace formula for maps (21.9):

return to p. 556

$$\sum_{k=0}^{\infty} \frac{ze^{s_k}}{1 - ze^{s_k}} = \sum_{r=1}^{\infty} \frac{z^r}{|1 - \Lambda^r|}, \quad e^{s_k} = \frac{1}{|\Lambda|\Lambda^k}. \quad (28.22)$$

Example 28.4. Meromorphic functions and exponential convergence: As an illustration of how exponential convergence of a truncated series is related to analytic properties of functions, consider, as the simplest possible example of a meromorphic function, the ratio

$$h(z) = \frac{z - a}{z - b}$$

with a, b real and positive and $a < b$. Within the *spectral radius* $|z| < b$ the function h can be represented by the power series

$$h(z) = \sum_{k=0}^{\infty} \sigma_k z^k,$$

where $\sigma_0 = a/b$, and the higher order coefficients are given by $\sigma_j = (a-b)/b^{j+1}$. Consider now the truncation of order N of the power series

return to p. 556

$$h_N(z) = \sum_{k=0}^N \sigma_k z^k = \frac{a}{b} + \frac{z(a-b)(1 - z^N/b^N)}{b^2(1 - z/b)}.$$

Let $\hat{z}_N$ be the solution of the truncated series $h_N(\hat{z}_N) = 0$. To estimate the distance between a and $\hat{z}_N$ it is sufficient to calculate $h_N(a)$. It is of order $(a/b)^{N+1}$, so finite order estimates converge exponentially to the asymptotic value.

Example 28.5. The spectral determinant for a single fixed point: The spectral determinant (22.3) follows from the trace formulas of example 28.3:

$$\det(1 - z\mathcal{L}) = \prod_{k=0}^{\infty} \left(1 - \frac{z}{|\Lambda|\Lambda^k}\right) = \sum_{n=0}^{\infty} (-t)^n Q_n, \quad t = \frac{z}{|\Lambda|}, \quad (28.23)$$

where the cummulants Q_n are given explicitly by the *Euler formula*

exercise 28.3
return to p. 556

$$Q_n = \frac{1}{1 - \Lambda^{-1}} \frac{\Lambda^{-1}}{1 - \Lambda^{-2}} \cdots \frac{\Lambda^{-n+1}}{1 - \Lambda^{-n}}. \quad (28.24)$$

Example 28.6. The simplest of 2-dimensional maps - a single hyperbolic fixed point: We start by considering a very simple linear hyperbolic map with a single hyperbolic fixed point,

$$f(x) = (f_1(x_1, x_2), f_2(x_1, x_2)) = (\Lambda_s x_1, \Lambda_u x_2), \quad 0 < |\Lambda_s| < 1, \quad |\Lambda_u| > 1.$$

The Perron-Frobenius operator (19.10) acts on the 2-dimensional density functions as

$$\mathcal{L}\rho(x_1, x_2) = \frac{1}{|\Lambda_s \Lambda_u|} \rho(x_1/\Lambda_s, x_2/\Lambda_u) \quad (28.25)$$

What are good eigenfunctions? Cribbing the 1-dimensional eigenfunctions for the stable, contracting x_1 direction from example 28.1 is not a good idea, as under the iteration of $\mathcal{L}$ the high terms in a Taylor expansion of $\rho(x_1, x_2)$ in the x_1 variable would get multiplied by exponentially exploding eigenvalues $1/\Lambda_s^k$. This makes sense, as in the contracting directions hyperbolic dynamics crunches up initial densities, instead of smoothing them. So we guess instead that the eigenfunctions are of form

$$\varphi_{k_1 k_2}(x_1, x_2) = x_2^{k_2} / x_1^{k_1+1}, \quad k_1, k_2 = 0, 1, 2, \dots, \quad (28.26)$$

a mixture of the Laurent series in the contraction x_1 direction, and the Taylor series in the expanding direction, the x_2 variable. The action of Perron-Frobenius operator on this set of basis functions

$$\mathcal{L}\varphi_{k_1 k_2}(x_1, x_2) = \frac{\sigma}{|\Lambda_u|} \frac{\Lambda_s^{k_1}}{\Lambda_u^{k_2}} \varphi_{k_1 k_2}(x_1, x_2), \quad \sigma = \Lambda_s / |\Lambda_s|$$

is smoothing, with the higher k_1, k_2 eigenvectors decaying exponentially faster, by $\Lambda_s^{k_1} / \Lambda_u^{k_2+1}$ factor in the eigenvalue. One verifies by an explicit calculation (undoing the geometric series expansions to lead to (22.8)) that the trace of $\mathcal{L}$ indeed equals $1/|\det(\mathbf{1} - M)| = 1/|(1 - \Lambda_u)(1 - \Lambda_s)|$, from which it follows that all our trace and spectral determinant formulas apply. The argument applies to any hyperbolic map linearized around the fixed point of form $f(x_1, \dots, x_d) = (\Lambda_1 x_1, \Lambda_2 x_2, \dots, \Lambda_d x_d)$.

return to p. 557

Example 28.7. Perron-Frobenius operator in a matrix representation: As in example 28.1, we start with a map with a single fixed point, but this time with a nonlinear piecewise analytic map f with a nonlinear inverse $F = f^{-1}$, sign of the derivative $\sigma = \sigma(F') = F'/|F'|$, and the Perron-Frobenius operator acting on densities analytic in an open domain enclosing the fixed point $x = w_q$,

$$\mathcal{L}\phi(y) = \int dx \delta(y - f(x)) \phi(x) = \sigma F'(y) \phi(F(y)).$$

Assume that F is a contraction of the unit disk in the complex plane, i.e.,

$$|F(z)| < \theta < 1 \quad \text{and} \quad |F'(z)| < C < \infty \quad \text{for} \quad |z| < 1, \quad (28.27)$$

and expand ϕ in a polynomial basis with the Cauchy integral formula

$$\phi(z) = \sum_{n=0}^{\infty} z^n \phi_n = \oint \frac{dw}{2\pi i} \frac{\phi(w)}{w - z}, \quad \phi_n = \oint \frac{dw}{2\pi i} \frac{\phi(w)}{w^{n+1}}$$

Combining this with (28.10), we see that in this basis Perron-Frobenius operator $\mathcal{L}$ is represented by the matrix

$$\mathcal{L}\phi(w) = \sum_{m,n} w^m \mathbf{L}_{mn} \phi_n, \quad \mathbf{L}_{mn} = \oint \frac{dw}{2\pi i} \frac{\sigma F'(w)(F(w))^n}{w^{m+1}}. \quad (28.28)$$

Taking the trace and summing we get:

$$\text{Tr } \mathcal{L} = \sum_{n \geq 0} \mathbf{L}_{nn} = \oint \frac{dw}{2\pi i} \frac{\sigma F'(w)}{w - F(w)}.$$

This integral has but one simple pole at the unique fixed point $w^* = F(w^*) = f(w^*)$. Hence

$$\text{Tr } \mathcal{L} = \frac{\sigma F'(w^*)}{1 - F'(w^*)} = \frac{1}{|f'(w^*) - 1|}.$$

exercise 28.6
return to p. 559

Example 28.8. Bernoulli shift eigenfunctions: Let us revisit the Bernoulli shift example (28.19) on the space of analytic functions on a disk: apart from the origin we have only simple eigenvalues $\lambda_k = 2^{-k}$, $k = 0, 1, \dots$. The eigenvalue $\lambda_0 = 1$ corresponds to probability conservation: the corresponding eigenfunction $B_0(x) = 1$ indicates that the natural measure has a constant density over the unit interval. If we now take any analytic function $\eta(x)$ with zero average (with respect to the Lebesgue measure), it follows that $\omega_1(\eta, \eta) = 0$, and from (28.16) the asymptotic decay of the correlation function is (unless also $\omega_1(\eta, \eta) = 0$)

$$C_{\eta, \eta}(n) \sim \exp(-n \log 2). \quad (28.29)$$

Thus, $-\log \lambda_1$ gives the exponential decay rate of correlations (with a prefactor that depends on the choice of the function). Actually the Bernoulli shift case may be treated exactly, as for analytic functions we can employ the Euler-MacLaurin summation formula

$$\eta(z) = \int_0^1 dw \eta(w) + \sum_{m=1}^{\infty} \frac{\eta^{(m-1)}(1) - \eta^{(m-1)}(0)}{m!} B_m(z). \quad (28.30)$$

As we are considering functions with zero average, we have from (28.15) and the fact that Bernoulli polynomials are eigenvectors of the Perron-Frobenius operator that

$$C_{\eta, \eta}(n) = \sum_{m=1}^{\infty} \frac{(2^{-m})^n (\eta^{(m)}(1) - \eta^{(m)}(0))}{m!} \int_0^1 dz \eta(z) B_m(z).$$

The decomposition (28.30) is also useful in realizing that the linear functionals ψ_i^* are singular objects: if we write it as

$$\eta(z) = \sum_{m=0}^{\infty} B_m(z) \psi_m^*[\eta],$$

we see that these functionals are of the form

$$\psi_i^*[\varepsilon] = \int_0^1 dw \Psi_i(w) \varepsilon(w),$$

where

$$\Psi_i(w) = \frac{(-1)^{i-1}}{i!} \left(\delta^{(i-1)}(w-1) - \delta^{(i-1)}(w) \right), \quad (28.31)$$

when $i \geq 1$ and $\Psi_0(w) = 1$. This representation is only meaningful when the function ε is analytic in neighborhoods of $w, w-1$. (continued in example 29.1)

[return to p. 569](#)

28.9 Exercises

- 28.1. **What space does $\mathcal{L}$ act on?** Show that (28.18) is a complete basis on the space of analytic functions on a disk (and thus that we found the *complete* set of eigenvalues).
- 28.2. **What space does $\mathcal{L}$ act on?** What can be said about the spectrum of (28.17) on $L^1[0, 1]$? Compare the result with figure 28.5.
- 28.3. **Euler formula.** Derive the Euler formula (28.24), $|u| < 1$:

$$\begin{aligned} \prod_{k=0}^{\infty} (1 + tu^k) &= 1 + \frac{t}{1-u} + \frac{t^2 u}{(1-u)(1-u^2)} \\ &\quad + \frac{t^3 u^3}{(1-u)(1-u^2)(1-u^3)} \cdots \\ &= \sum_{k=0}^{\infty} t^k \frac{u^{\frac{k(k-1)}{2}}}{(1-u) \cdots (1-u^k)}. \end{aligned}$$

- 28.4. **2-dimensional product expansion.**  We conjecture that the expansion corresponding to exercise 28.3 is in the 2-dimensional case given by

$$\begin{aligned} &\prod_{k=0}^{\infty} (1 + tu^k)^{k+1} \\ &= \sum_{k=0}^{\infty} \frac{F_k(u)}{(1-u)^2(1-u^2)^2 \cdots (1-u^k)^2} t^k \\ &= 1 + \frac{1}{(1-u)^2} t + \frac{2u}{(1-u)^2(1-u^2)^2} t^2 \\ &\quad + \frac{u^2(1+4u+u^2)}{(1-u)^2(1-u^2)^2(1-u^3)^2} t^3 + \cdots \end{aligned}$$

$F_k(u)$ is a polynomial in u , and the coefficients fall off asymptotically as $C_n \approx u^{n^{3/2}}$. Verify; if you have a proof to all orders, e-mail it to the authors. (See also solution 28.3).

- 28.5. **Bernoulli shift on L spaces.** Check that the family (28.9) belongs to $L^1([0, 1])$. What can be said about the essential spectral radius on $L^2([0, 1])$? A useful reference is ref. [26].
- 28.6. **Cauchy integrals.** Rework all complex analysis steps used in the Bernoulli shift example on analytic functions on a disk.
- 28.7. **Escape rate.** Consider the escape rate from a strange repeller: find a choice of trial functions ξ and φ such that (28.15) gives the fraction on particles surviving after n iterations, if their initial density distribution is $\rho_0(x)$. Discuss the behavior of such an expression in the long time limit.

Chapter 29

Intermittency

Sometimes They Come Back

—Stephen King

(R. Artuso, P. Dahlqvist, G. Tanner and P. Cvitanović)

IN THE THEORY of chaotic dynamics developed so far we assumed that the evolution operators have discrete spectra $\{z_0, z_1, z_2, \dots\}$ given by the zeros of

$$1/\zeta(z) = (\cdots) \prod_k (1 - z/z_k).$$

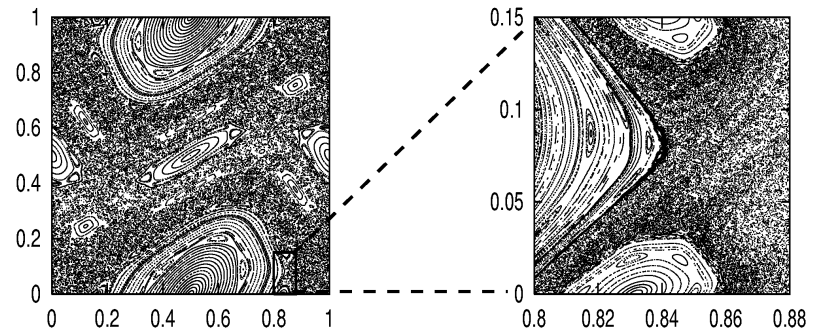
The assumption was based on the tacit premise that the dynamics is everywhere exponentially unstable. Real life is nothing like that - state spaces are generically infinitely interwoven patterns of stable and unstable behaviors. The stable (in the case of Hamiltonian flows, integrable) orbits do not communicate with the ergodic components of the state space, and can be treated by deterministic methods. In general, one is able to treat the dynamics near stable orbits as well as chaotic components of the state space dynamics well within a periodic orbit approach. Problems occur at the borderline between chaos and regular dynamics where marginally stable orbits and manifolds present difficulties and still unresolved challenges.

We shall use the simplest example of such behavior - intermittency in 1-dimensional maps - to illustrate effects of marginal stability. The main message will be that spectra of evolution operators are no longer discrete, dynamical zeta functions exhibit branch cuts of the form

$$1/\zeta(z) = (\cdots) + (1 - z)^\alpha (\cdots),$$

and correlations decay no longer exponentially, but as power laws.

Figure 29.1: Typical phase space for an area-preserving map with mixed phase space dynamics; here the standard map for $k = 1.2$.



29.1 Intermittency everywhere

In many fluid dynamics experiments one observes transitions from regular behaviors to behaviors where long time intervals of regular behavior (“laminar phases”) are interrupted by fast irregular bursts. The closer the parameter is to the onset of such bursts, the longer are the intervals of regular behavior. The distributions of laminar phase intervals are well described by power laws.

This phenomenon is called *intermittency*, and it is a very general aspect of dynamics, a shadow cast by nonhyperbolic, marginally stable state space regions. Complete hyperbolicity assumed in (21.5) is the exception rather than the rule, and for almost any dynamical system of interest (dynamics in smooth potentials, billiards with smooth walls, the infinite horizon Lorentz gas, etc.) one encounters mixed state spaces with islands of stability coexisting with hyperbolic regions, see figure 29.1 and example 8.7. Wherever stable islands are interspersed with chaotic regions, trajectories which come close to the stable islands can stay ‘glued’ for arbitrarily long times. These intervals of regular motion are interrupted by irregular bursts as the trajectory is re-injected into the chaotic part of the state space. How the trajectories are precisely ‘glued’ to the marginally stable region is often hard to describe. What coarsely looks like a border of an island will under magnification dissolve into infinities of island chains of decreasing sizes, broken tori and bifurcating orbits, as illustrated in figure 29.1.

Intermittency is due to the existence of fixed points and cycles of marginal stability (5.6), or (in studies of the onset of intermittency) to the proximity of a nearly marginal complex or unstable orbits. In Hamiltonian systems intermittency goes hand in hand with the existence of (marginally stable) KAM tori. In more general settings, the existence of marginal or nearly marginal orbits is due to incomplete intersections of stable and unstable manifolds in a Smale horseshoe type dynamics (see figure 15.11). Following the stretching and folding of the invariant manifolds in time one will inevitably find state space points at which the stable and unstable manifolds are almost or exactly tangential to each other, implying non-exponential separation of nearby points in state space or, in other words, marginal stability. Under small parameter perturbations such neighborhoods undergo tangent bifurcations - a stable/unstable pair of periodic orbits is destroyed or created by coalescing into a marginal orbit, so the pruning which we shall encounter in chapter 15, and the intermittency discussed here are two sides of the same coin.

section 15.4

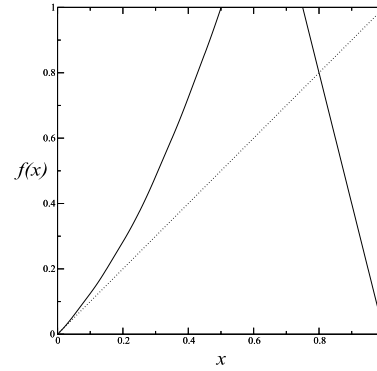


Figure 29.2: A complete binary repeller with a marginal fixed point.

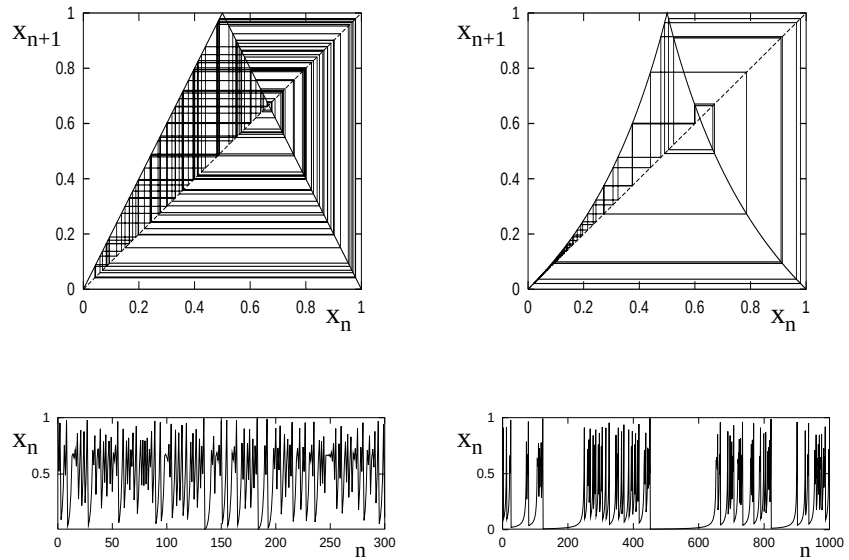


Figure 29.3: (a) A tent map trajectory. (b) A Farey map trajectory.

How to deal with the full complexity of a typical Hamiltonian system with mixed phase space is a very difficult, still open problem. Nevertheless, it is possible to learn quite a bit about intermittency by considering rather simple examples. Here we shall restrict our considerations to 1-dimensional maps which in the neighborhood of a single marginally stable fixed point at $x=0$ take the form

$$x \mapsto f(x) = x + O(x^{1+s}), \quad (29.1)$$

and are expanding everywhere else. Such a map may allow for escape, like the map shown in figure 29.2 or the dynamics may be bounded, like the Farey map

$$x \mapsto f(x) = \begin{cases} x/(1-x) & x \in [0, 1/2[\\ (1-x)/x & x \in [1/2, 1] \end{cases} \quad (29.2)$$

Figure 29.3 compares a trajectory of the tent map (14.21) side by side with a trajectory of the Farey map. In a stark contrast to the uniformly chaotic trajectory of the tent map, the Farey map trajectory alternates intermittently between slow regular motion close to the marginally stable fixed point, and chaotic bursts.

section 29.3.4

The presence of marginal stability has striking dynamical consequences: correlation decay may exhibit long range power law asymptotic behavior and diffusion processes can assume anomalous character. Escape from a repeller of the

form figure 29.2 may be algebraic rather than exponential. In long time explorations of the dynamics intermittency manifests itself by enhancement of natural measure in the proximity of marginally stable cycles.

The questions we shall address here are: how does marginal stability affect zeta functions or spectral determinants? And, can we deduce power law decays of correlations from cycle expansions?

In example 28.2 we saw that marginal stability violates one of the conditions which ensure that the spectral determinant is an entire function. Already the simple fact that the cycle weight $1/|1 - \Lambda_p^r|$ in the trace (21.3) or the spectral determinant (22.3) diverges for marginal orbits with $|\Lambda_p| = 1$ tells us that we have to treat these orbits with care.

In the following we will incorporate marginal stability orbits into cycle-expansions in a systematic manner. To get to know the difficulties lying ahead, we will start in sect. 29.2 with a piecewise linear map, with the asymptotics (29.1). We will construct a dynamical zeta function in the usual way without worrying too much about its justification and show that it has a branch cut singularity. We will calculate the rate of escape from our piecewise linear map and find that it is characterized by decay, rather than exponential decay, a power law. We will show that dynamical zeta functions in the presence of marginal stability can still be written in terms of periodic orbits, exactly as in chapters 20 and 27, with one exception: the marginally stable orbits have to be explicitly excluded. This innocent looking step has far reaching consequences; it forces us to change the symbolic dynamics from a finite to an infinite alphabet, and entails a reorganization of the order of summations in cycle expansions, sect. 29.2.4.

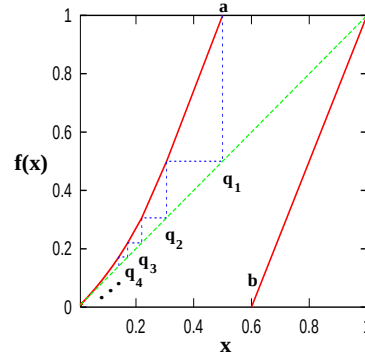
Branch cuts are typical also for smooth intermittent maps with isolated marginally stable fixed points and cycles. In sect. 29.3, we discuss the cycle expansions and curvature combinations for zeta functions of smooth maps tailored to intermittency. The knowledge of the type of singularity one encounters enables us to develop the efficient resummation method presented in sect. 29.3.1.

Finally, in sect. 29.4, we discuss a probabilistic approach to intermittency that yields approximate dynamical zeta functions and provides valuable information about more complicated systems, such as billiards.

29.2 Intermittency for pedestrians

Intermittency does not only present us with a large repertoire of interesting dynamics, it is also at the root of many sorrows such as slow convergence of cycle expansions. In order to get to know the kind of problems which arise when studying dynamical zeta functions in the presence of marginal stability we will consider an artfully concocted piecewise linear model first. From there we will move on to the more general case of smooth intermittant maps, sect. 29.3.

Figure 29.4: A piecewise linear intermittent map of (29.3) type: more specifically, the map piecewise linear over intervals (29.8) of the toy example studied below, $a = .5$, $b = .6$, $s = 1.0$.



29.2.1 A toy map

The Bernoulli shift map (28.19) is an idealized, but highly instructive, example of a hyperbolic map. To study intermittency we will now construct a likewise piecewise linear model, an intermittent map stripped down to its bare essentials.

Consider a map $x \mapsto f(x)$ on the unit interval $\mathcal{M} = [0, 1]$ with two monotone branches

$$f(x) = \begin{cases} f_0(x) & \text{for } x \in \mathcal{M}_0 = [0, a] \\ f_1(x) & \text{for } x \in \mathcal{M}_1 = [b, 1] \end{cases} \quad (29.3)$$

The two branches are assumed complete, that is $f_0(\mathcal{M}_0) = f_1(\mathcal{M}_1) = \mathcal{M}$. The map allows escape if $a < b$ and is bounded if $a = b$ (see figure 29.2 and figure 29.4). We take the right branch to be expanding and linear:

$$f_1(x) = \frac{1}{1-b}(x-b).$$

Next, we will construct the left branch in a way, which will allow us to model the intermittent behavior (29.1) near the origin. We chose a monotonically decreasing sequence of points q_n in $[0, a]$ with $q_1 = a$ and $q_n \rightarrow 0$ as $n \rightarrow \infty$. This sequence defines a partition of the left interval $\mathcal{M}_0$ into an infinite number of connected intervals $\mathcal{M}_n$, $n \geq 2$ with

$$\mathcal{M}_n =]q_n, q_{n-1}] \quad \text{and} \quad \mathcal{M}_0 = \bigcup_{n=2}^{\infty} \mathcal{M}_n. \quad (29.4)$$

The map $f_0(x)$ is now specified by the following requirements

- $f_0(x)$ is continuous.
- $f_0(x)$ is linear on the intervals $\mathcal{M}_n$ for $n \geq 2$.
- $f_0(q_n) = q_{n-1}$, that is $\mathcal{M}_n = f_0^{-n+1}([a, 1])$.

This fixes the map for any given sequence $\{q_n\}$. The last condition ensures the existence of a simple Markov partition. The slopes of the various linear segments

are

$$\begin{aligned} f'_0(x) &= \frac{f_0(q_{n-1}) - f_0(q_n)}{q_{n-1} - q_n} = \frac{|\mathcal{M}_{n-1}|}{|\mathcal{M}_n|} & \text{for } x \in \mathcal{M}_n, n \geq 3 \\ f'_0(x) &= \frac{f_0(q_1) - f_0(q_2)}{q_1 - q_2} = \frac{1-a}{|\mathcal{M}_2|} & \text{for } x \in \mathcal{M}_2 \\ f'_0(x) &= \frac{1}{1-b} = \frac{|\mathcal{M}|}{|\mathcal{M}_1|} & \text{for } x \in \mathcal{M}_1 \end{aligned} \quad (29.5)$$

with $|\mathcal{M}_n| = q_{n-1} - q_n$ for $n \geq 2$. Note that we do not require as yet that the map exhibit intermittent behavior.

We will see that the family of periodic orbits with code 10^n plays a key role for intermittent maps of the form (29.1). An orbit 10^n enters the intervals $\mathcal{M}_1 \rightarrow \mathcal{M}_{n+1} \rightarrow \mathcal{M}_n \rightarrow \dots \rightarrow \mathcal{M}_2$ successively and the family approaches the marginal stable fixed point at $x = 0$ for $n \rightarrow \infty$. The stability of a cycle 10^n for $n \geq 1$ is given by the chain rule (4.43),

$$\Lambda_{10^n} = f'_0(x_{n+1})f'_0(x_n) \dots f'_0(x_2)f'_1(x_1) = \frac{1}{|\mathcal{M}_{n+1}|} \frac{1-a}{1-b}, \quad (29.6)$$

with $x_i \in \mathcal{M}_i$.

The properties of the map (29.3) are completely determined by the sequence $\{q_n\}$. By choosing $q_n = 2^{-n}$, for example, we recover the uniformly hyperbolic Bernoulli shift map (28.19). An intermittent map of the form (29.4) having the asymptotic behavior (29.1) can be constructed by choosing an algebraically decaying sequence $\{q_n\}$ behaving asymptotically like

$$q_n \sim \frac{1}{n^{1/s}},$$

where s is the intermittency exponent in (29.1). Such a partition leads to intervals whose length decreases asymptotically like a power-law, that is,

$$|\mathcal{M}_n| \sim \frac{1}{n^{1+1/s}}. \quad (29.7)$$

As can be seen from (29.6), the Floquet multipliers of periodic orbit families approaching the marginal fixed point, such as the 10^n family increase in turn only algebraically with the cycle period.

It may now seem natural to construct an intermittent toy map in terms of a partition $|\mathcal{M}_n| = 1/n^{1+1/s}$, that is, a partition which follows (29.7) exactly. Such a choice leads to a dynamical zeta function which can be written in terms of so-called Jonquière functions (or polylogarithms) which arise naturally also in the context of the Farey map (29.2), and the anomalous diffusion of sect. 24.3. We will, however, not follow this route here; instead, we will engage in a bit of reverse engineering and construct a less obvious partition which will simplify the algebra considerably later without loosing any of the key features typical for intermittent systems. We fix the intermittent toy map by specifying the intervals $\mathcal{M}_n$ in terms of Gamma functions according to

$$|\mathcal{M}_n| = C \frac{\Gamma(n+m-1/s-1)}{\Gamma(n+m)} \quad \text{for } n \geq 2, \quad (29.8)$$

question 24.1

where $m = [1/s]$ denotes the integer part of $1/s$ and C is a normalization constant fixed by the condition $\sum_{n=2}^{\infty} |\mathcal{M}_n| = q_1 = a$, that is,

$$C = a \left[\sum_{n=m+1}^{\infty} \frac{\Gamma(n - 1/s)}{\Gamma(n + 1)} \right]^{-1}. \quad (29.9)$$

Using Stirling's formula for the Gamma function

$$\Gamma(z) \sim e^{-z} z^{z-1/2} \sqrt{2\pi} (1 + 1/12z + \dots),$$

we verify that the intervals decay asymptotically like $n^{-(1+1/s)}$, as required by the condition (29.7).

Next, let us write down the dynamical zeta function of the toy map in terms of its periodic orbits, that is

$$1/\zeta(z) = \prod_p \left(1 - \frac{z^{n_p}}{|\Lambda_p|} \right)$$

One may be tempted to expand the dynamical zeta function in terms of the binary symbolic dynamics of the map; we saw, however, in sect. 23.7 that such cycle expansion converges extremely slowly. The shadowing mechanism between orbits and pseudo-orbits fails for orbits of the form 10^n with stabilities given by (29.6), due to the marginal stability of the fixed point $\bar{0}$. It is therefore advantageous to choose as the fundamental cycles the family of orbits with code 10^n or, equivalently, switch from the finite (binary) alphabet to an infinite alphabet given by

$$10^{n-1} \rightarrow n.$$

Due to the piecewise-linear form of the map which maps intervals $\mathcal{M}_n$ exactly onto $\mathcal{M}_{n-1}$, all periodic orbits entering the left branch at least twice are canceled exactly by pseudo cycles, and the cycle expanded dynamical zeta function depends only on the fundamental series $1, 10, 100, \dots$:

$$\begin{aligned} 1/\zeta(z) &= \prod_{p \neq 0} \left(1 - \frac{z^{n_p}}{|\Lambda_p|} \right) = 1 - \sum_{n=1}^{\infty} \frac{z^n}{|\Lambda_{10^{n-1}}|} \\ &= 1 - (1-b)z - C \frac{1-b}{1-a} \sum_{n=2}^{\infty} \frac{\Gamma(n+m-1/s-1)}{\Gamma(n+m)} z^n. \end{aligned} \quad (29.10)$$

The fundamental term (23.8) consists here of an infinite sum over algebraically decaying cycle weights. The sum is divergent for $|z| \geq 1$. We will see that this behavior is due to a branch cut of $1/\zeta$ starting at $z = 1$. We need to find analytic continuations of sums over algebraically decreasing terms in (29.10). Note also that we omitted the fixed point $\bar{0}$ in the above Euler product; we will discuss this point as well as a proper derivation of the zeta function in more detail in sect. 29.2.4.

29.2.2 Branch cuts

Starting from the dynamical zeta function (29.10), we first have to worry about finding an analytical continuation of the sum for $|z| \geq 1$. We do, however, get this part for free here due to the particular choice of interval lengths made in (29.8). The sum over ratios of Gamma functions in (29.10) can be evaluated analytically by using the following identities valid for $1/s = \alpha > 0$ (the famed binomial theorem in disguise),

- α non-integer

$$(1-z)^\alpha = \sum_{n=0}^{\infty} \frac{\Gamma(n-\alpha)}{\Gamma(-\alpha)\Gamma(n+1)} z^n \quad (29.11)$$

- α integer

$$\begin{aligned} (1-z)^\alpha \log(1-z) &= \sum_{n=1}^{\alpha} (-1)^n c_n z^n \\ &+ (-1)^{\alpha+1} \alpha! \sum_{n=\alpha+1}^{\infty} \frac{(n-\alpha-1)!}{n!} z^n \end{aligned} \quad (29.12)$$

with

$$c_n = \binom{\alpha}{n} \sum_{k=0}^{n-1} \frac{1}{\alpha-k}.$$

In order to simplify the notation, we restrict the intermittency parameter to the range $1 \leq 1/s < 2$ with $[1/s] = m = 1$. All what follows can easily be generalized to arbitrary $s > 0$ using equations (29.11) and (29.12). The infinite sum in (29.10) can now be evaluated with the help of (29.11) or (29.12), that is,

$$\sum_{n=2}^{\infty} \frac{\Gamma(n-1/s)}{\Gamma(n+1)} z^n = \begin{cases} \Gamma(-\frac{1}{s}) \left[(1-z)^{1/s} - 1 + \frac{1}{s} z \right] & \text{for } 1 < 1/s < 2; \\ (1-z) \log(1-z) + z & \text{for } s = 1. \end{cases}$$

The normalization constant C in (29.8) can be evaluated explicitly using (29.9) and the dynamical zeta function can be given in closed form. We obtain for $1 < 1/s < 2$

$$1/\zeta(z) = 1 - (1-b)z - \frac{a}{1/s-1} \frac{1-b}{1-a} \left((1-z)^{1/s} - 1 + \frac{1}{s} z \right). \quad (29.13)$$

and for $s = 1$,

$$1/\zeta(z) = 1 - (1-b)z - a \frac{1-b}{1-a} ((1-z) \log(1-z) + z). \quad (29.14)$$

It now becomes clear why the particular choice of intervals $\mathcal{M}_n$ made in the last section is useful; by summing over the infinite family of periodic orbits $O^n 1$ explicitly, we have found the desired analytical continuation for the dynamical zeta

function for $|z| \geq 1$. The function has a branch cut starting at the branch point $z = 1$ and running along the positive real axis. That means, the dynamical zeta function takes on different values when approaching the positive real axis for $\operatorname{Re} z > 1$ from above and below. The dynamical zeta function for general $s > 0$ takes on the form

$$1/\zeta(z) = 1 - (1-b)z - \frac{a}{g_s(1)} \frac{1-b}{1-a} \frac{1}{z^{m-1}} \left((1-z)^{1/s} - g_s(z) \right) \quad (29.15)$$

for non-integer s with $m = [1/s]$ and

$$1/\zeta(z) = 1 - (1-b)z - \frac{a}{g_m(1)} \frac{1-b}{1-a} \frac{1}{z^{m-1}} \left((1-z)^m \log(1-z) - g_m(z) \right) \quad (29.16)$$

for $1/s = m$ integer and $g_s(z)$ are polynomials of order $m = [1/s]$ which can be deduced from (29.11) or (29.12). We thus find algebraic branch cuts for non integer intermittency exponents $1/s$ and logarithmic branch cuts for $1/s$ integer. We will see in sect. 29.3 that branch cuts of that form are generic for 1-dimensional intermittent maps.

Branch cuts are the all important new feature of dynamical zeta functions due to intermittency. So, how do we calculate averages or escape rates of the dynamics of the map from a dynamical zeta function with branch cuts? We take ‘a learning by doing’ approach and calculate the escape from our toy map for $a < b$.

29.2.3 Escape rate

Our starting point for the calculation of the fraction of survivors after n time steps, is the integral representation (A22.6)

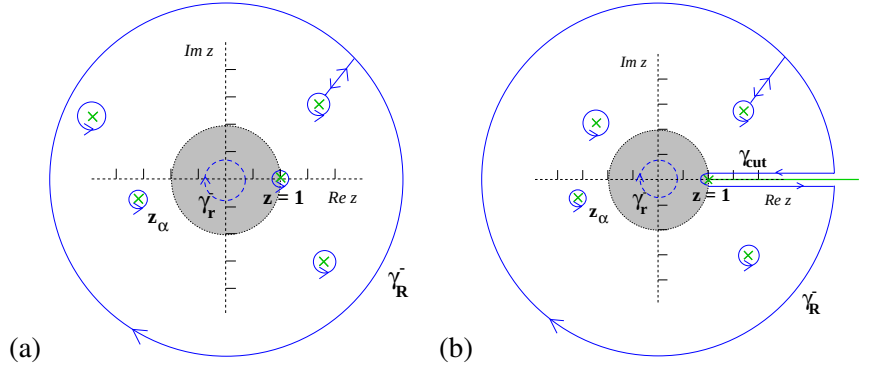
$$\Gamma_n = \frac{1}{2\pi i} \oint_{\gamma_r^-} z^{-n} \left(\frac{d}{dz} \log \zeta^{-1}(z) \right) dz, \quad (29.17)$$

where the contour encircles the origin in the clockwise direction. If the contour lies inside the unit circle $|z| = 1$, we may expand the logarithmic derivative of $\zeta^{-1}(z)$ as a convergent sum over all periodic orbits. Integrals and sums can be interchanged, the integrals can be solved term by term, and the formula (21.22) is recovered. For hyperbolic maps, cycle expansion methods or other techniques may provide an analytic extension of the dynamical zeta function beyond the leading zero; we may therefore deform the original contour into a larger circle with radius R which encircles both poles and zeros of $\zeta^{-1}(z)$, see figure 29.5 (a). Residue calculus turns this into a sum over the zeros z_α and poles z_β of the dynamical zeta function, that is

$$\Gamma_n = \sum_{|z_\alpha| < R}^{\text{zeros}} \frac{1}{z_\alpha^n} - \sum_{|z_\beta| < R}^{\text{poles}} \frac{1}{z_\beta^n} + \frac{1}{2\pi i} \oint_{\gamma_R^-} dz z^{-n} \frac{d}{dz} \log \zeta^{-1}, \quad (29.18)$$

where the last term gives a contribution from a large circle γ_R^- . We thus find exponential decay of Γ_n dominated by the leading zero or pole of $\zeta^{-1}(z)$.

Figure 29.5: The survival probability Γ_n calculated by contour integration; integrating (29.17) inside the domain of convergence $|z| < 1$ (shaded area) of $1/\zeta(z)$ in periodic orbit representation yields (21.22). A deformation of the contour γ_r^- (dashed circle) to a larger circle γ_R^- gives contributions from the poles and zeros $\otimes$ of $1/\zeta(z)$ between the two circles. These are the only contributions for hyperbolic maps (a), for intermittent systems additional contributions arise, given by the contour γ_{cut} running along the branch cut (b).



Things change considerably in the intermittent case. The point $z = 1$ is a branch cut singularity and there exists no Taylor series expansion of ζ^{-1} around $z = 1$. Second, the path deformation that led us to (29.18) requires more care, as it must not cross the branch cut. When expanding the contour to large $|z|$ values, we have to deform it along the branch $\text{Re}(z) \geq 1, \text{Im}(z) = 0$ encircling the branch cut in anti-clockwise direction, see figure 29.5 (b). We will denote the detour around the cut as γ_{cut} . We may write symbolically

$$\oint_{\gamma_r} = \sum_{\text{zeros}} - \sum_{\text{poles}} + \oint_{\gamma_R} + \oint_{\gamma_{cut}}$$

where the sums include only the zeros and the poles in the area enclosed by the contours. The asymptotics is controlled by the zero, pole or cut closest to the origin.

Let us now go back to our intermittent toy map. The asymptotics of the survival probability of the map is here governed by the behavior of the integrand $\frac{d}{dz} \log \zeta^{-1}$ in (29.17) at the branch point $z = 1$. We restrict ourselves again to the case $1 < 1/s < 2$ first and write the dynamical zeta function (29.13) in the form

$$1/\zeta(z) = a_0 + a_1(1-z) + b_0(1-z)^{1/s} \equiv G(1-z)$$

and

$$a_0 = \frac{b-a}{1-a}, \quad b_0 = \frac{a}{1-1/s} \frac{1-b}{1-a}.$$

Setting $u = 1 - z$, we need to evaluate

$$\frac{1}{2\pi i} \oint_{\gamma_{cut}} (1-u)^{-n} \frac{d}{du} \log G(u) du \quad (29.19)$$

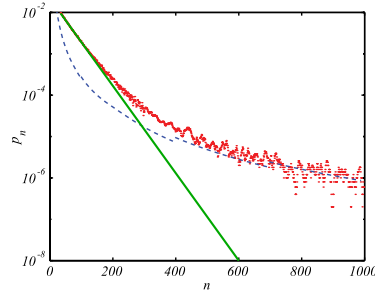
where γ_{cut} goes around the cut (i.e., the negative u axis). Expanding the integrand $\frac{d}{du} \log G(u) = G'(u)/G(u)$ in powers of u and $u^{1/s}$ at $u = 0$, one obtains

$$\frac{d}{du} \log G(u) = \frac{a_1}{a_0} + \frac{1}{s} \frac{b_0}{a_0} u^{1/s-1} + O(u). \quad (29.20)$$

The integrals along the cut may be evaluated using the general formula

$$\frac{1}{2\pi i} \oint_{\gamma_{cut}} u^\alpha (1-u)^{-n} du = \frac{\Gamma(n-\alpha-1)}{\Gamma(n)\Gamma(-\alpha)} \sim \frac{1}{n^{\alpha+1}} (1 + O(1/n)) \quad (29.21)$$

Figure 29.6: The asymptotic escape from an intermittent repeller is a power law. Normally it is preceded by an exponential, which can be related to zeros close to the cut but beyond the branch point $z = 1$, as in figure 29.5 (b).



which can be obtained by deforming the contour back to a loop around the point $u = 1$, now in positive (anti-clockwise) direction. The contour integral then picks up the $(n-1)$ st term in the Taylor expansion of the function u^α at $u = 1$, cf. (29.11). For the continuous time case the corresponding formula is

$$\frac{1}{2\pi i} \oint_{\gamma_{cut}} z^\alpha e^{zt} dz = \frac{1}{\Gamma(-\alpha)} \frac{1}{t^{\alpha+1}}. \quad (29.22)$$

Plugging (29.20) into (29.19) and using (29.21) we get the asymptotic result

$$\Gamma_n \sim \frac{b_0}{a_0} \frac{1}{s} \frac{1}{\Gamma(1-1/s)} \frac{1}{n^{1/s}} = \frac{a}{s-1} \frac{1-b}{b-a} \frac{1}{\Gamma(1-1/s)} \frac{1}{n^{1/s}}. \quad (29.23)$$

We see that, asymptotically, the escape from an intermittent repeller is described by power law decay rather than the exponential decay we are familiar with for hyperbolic maps; a numerical simulation of the power-law escape from an intermittent repeller is shown in figure 29.6.

For general non-integer $1/s > 0$, we write

$$1/\zeta(z) = A(u) + (u)^{1/s} B(u) \equiv G(u)$$

with $u = 1 - z$ and $A(u)$, $B(u)$ are functions analytic in a disc of radius 1 around $u = 0$. The leading terms in the Taylor series expansions of $A(u)$ and $B(u)$ are

$$a_0 = \frac{b-a}{1-a}, \quad b_0 = \frac{a}{g_s(1)} \frac{1-b}{1-a},$$

see (29.15). Expanding $\frac{d}{du} \log G(u)$ around $u = 0$, one again obtains leading order contributions according to (29.20) and the general result follows immediately using (29.21), that is,

$$\Gamma_n \sim \frac{a}{s g_s(1)} \frac{1-b}{b-a} \frac{1}{\Gamma(1-1/s)} \frac{1}{n^{1/s}}. \quad (29.24)$$

Applying the same arguments for integer intermittency exponents $1/s = m$, one obtains

$$\Gamma_n \sim (-1)^{m+1} \frac{a}{s g_m(1)} \frac{1-b}{b-a} \frac{m!}{n^m}. \quad (29.25)$$

So far, we have considered the survival probability for a repeller, that is we assumed $a < b$. The formulas (29.24) and (29.25) do obviously not apply for the

case $a = b$, that is, for the bounded map. The coefficient $a_0 = (b - a)/(1 - a)$ in the series representation of $G(u)$ is zero, and the expansion of the logarithmic derivative of $G(u)$ (29.20) is no longer valid. We get instead

$$\frac{d}{du} \log G(u) = \begin{cases} \frac{1}{u} \left(1 + O(u^{1/s-1}) \right) & s < 1 \\ \frac{1}{u} \left(\frac{1}{s} + O(u^{1-1/s}) \right) & s > 1 \end{cases},$$

assuming non-integer $1/s$ for convenience. One obtains for the survival probability.

$$\Gamma_n \sim \begin{cases} 1 + O(n^{1-1/s}) & s < 1 \\ 1/s + O(n^{1/s-1}) & s > 1 \end{cases}.$$

For $s > 1$, this is what we expect. There is no escape, so the survival probability is equal to 1, which we get as an asymptotic result here. The result for $s > 1$ is somewhat more worrying. It says that Γ_n defined as sum over the instabilities of the periodic orbits as in (27.18) does not tend to unity for large n . However, the case $s > 1$ is in many senses anomalous. For instance, the invariant density cannot be normalized. It is therefore not reasonable to expect that periodic orbit theories will work without complications.

29.2.4 Why does it work (anyway)?

Due to the piecewise linear nature of the map constructed in the previous section, we had the nice property that interval lengths did exactly coincide with the inverse of the stability of periodic orbits of the system, that is

$$|\mathcal{M}_n| = 1/|\Lambda_{10}|^{n-1}.$$

There is thus no problem in replacing the survival probability Γ_n given by (1.2), (27.2), that is the fraction of state space $\mathcal{M}$ surviving n iterations of the map,

$$\Gamma_n = \frac{1}{|\mathcal{M}|} \sum_i^{(n)} |\mathcal{M}_i|.$$

by a sum over periodic orbits of the form (21.22). The only orbit to worry about is the marginal fixed point $\bar{0}$ itself which we excluded from the zeta function (29.10).

For smooth intermittent maps, things are less clear and the fact that we had to prune the marginal fixed point is a warning sign that interval estimates by periodic orbit stabilities might go horribly wrong. The derivation of the survival probability in terms of cycle stabilities in chapter 27 did indeed rely heavily on a hyperbolicity assumption which is clearly not fulfilled for intermittent maps. We therefore have to carefully reconsider this derivation in order to show that periodic orbit formulas are actually valid for intermittent systems in the first place.

We will for simplicity consider maps, which have a finite number of say s branches defined on intervals $\mathcal{M}_s$ and we assume that the map maps each interval $\mathcal{M}_s$ onto $\mathcal{M}$, that is $f(\mathcal{M}_s) = \mathcal{M}$. This ensures the existence of a complete symbolic dynamics - just to make things easy (see figure 29.2).

The generating partition is composed of the domains $\mathcal{M}_s$. The n th level partition $\mathcal{C}^{(n)} = \{\mathcal{M}_i\}$ can be constructed iteratively. Here i 's are words $i = s_2 s_2 \dots s_n$ of length n , and the intervals $\mathcal{M}_i$ are constructed recursively

$$\mathcal{M}_{sj} = f_s^{-1}(\mathcal{M}_j), \quad (29.26)$$

where sj is the concatenation of letter s with word j of length $n_j < n$.

In what follows we will concentrate on the survival probability Γ_n , postponing other quantities of interest, such as averages, to later considerations. In establishing the equivalence of the survival probability and the periodic orbit formula for the escape rate for hyperbolic systems we have assumed that the map is expanding, with a minimal expansion rate $|f'(x)| \geq \Lambda_{\min} > 1$. This enabled us to bound the size of every survivor strip $\mathcal{M}_i$ by (27.6), the stability Λ_i of the periodic orbit i within the $\mathcal{M}_i$, and bound the survival probability by the periodic orbit sum (27.7).

The bound (27.6)

$$C_1 \frac{1}{|\Lambda_i|} < \frac{|\mathcal{M}_i|}{|\mathcal{M}|} < C_2 \frac{1}{|\Lambda_i|}$$

relies on hyperbolicity, and is thus indeed violated for intermittent systems. The problem is that now there is no lower bound on the expansion rate, the minimal expansion rate is $\Lambda_{\min} = 1$. The survivor strip $\mathcal{M}_{0^n}$ which includes the marginal fixed point is thus completely overestimated by $1/|\Lambda_{0^n}| = 1$ which is constant for all n .

exercise A22.1

However, bounding survival probability strip by strip is not what is required for establishing the bound (27.7). For intermittent systems a somewhat weaker bound can be established, saying that the average size of intervals *along a periodic orbit* can be bounded close to the stability of the periodic orbit for all but the interval $\mathcal{M}_{0^n}$. The weaker bound applies to averaging over each prime cycle p separately

$$C_1 \frac{1}{|\Lambda_p|} < \frac{1}{n_p} \sum_{i \in p} \frac{|\mathcal{M}_i|}{|\mathcal{M}|} < C_2 \frac{1}{|\Lambda_p|}, \quad (29.27)$$

where the word i represents a code of the periodic orbit p and all its cyclic permutations. It can be shown that one can find positive constants C_1, C_2 independent of p . Summing over all periodic orbits leads then again to (27.7).

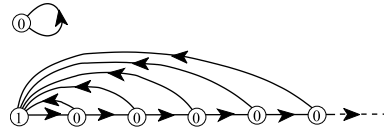
To study averages of multiplicative weights we follow sect. 20.1 and introduce a state space observable $a(x)$ and the integrated quantity

$$A(x, n) = \sum_{k=0}^{n-1} a(f^k(x)).$$

This leads us to introduce the moment-generating function (20.9)

$$\langle e^{\beta A(x, n)} \rangle,$$

Figure 29.7: Transition graph corresponding to the alphabet $\{0^{k-1}1; \bar{0}, k \geq 1\}$



where $\langle \cdot \rangle$ denote some averaging over the distribution of initial points, which we choose to be uniform (rather than the *a priori* unknown invariant density). Again, all we have to show is, that constants C_1, C_2 exist, such that

$$C_1 \frac{e^{\beta A_p}}{|\Lambda_p|} < \frac{1}{n_p} \sum_{i \in p} \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}_Q} e^{\beta A(x,n)} dx < C_2 \frac{e^{\beta A_p}}{|\Lambda_p|}, \quad (29.28)$$

is valid for all p . After performing the above average one gets

$$C_1 \Gamma_n(\beta) < \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} e^{\beta A(x,n)} dx < C_2 \Gamma_n(\beta), \quad (29.29)$$

with

$$\Gamma_n(\beta) = \sum_p \frac{e^{\beta A_p}}{|\Lambda_p|}.$$

and a dynamical zeta function can be derived. In the intermittent case one can expect that the bound (29.28) holds using an averaging argument similar to the one discussed in (29.27). This justifies the use of dynamical zeta functions for intermittent systems.

One lesson we should have learned so far is that the natural alphabet to use is not $\{0, 1\}$ but rather the infinite alphabet $\{0^{k-1}1, \bar{0}; k \geq 1\}$. The symbol $\bar{0}$ occurs unaccompanied by any 1's only in the $\bar{0}$ marginal fixed point which is disconnected from the rest of the transition graph, see figure 29.7.

chapter 15

What happens if we remove a single prime cycle from a dynamical zeta function? In the hyperbolic case such a removal introduces a pole in the $1/\zeta$ and slows down the convergence of cycle expansions. The heuristic interpretation of such a pole is that for a subshift of finite type removal of a single prime cycle leads to unbalancing of cancellations within the infinity of shadowing pairs. Nevertheless, removal of a single prime cycle is an exponentially small perturbation of the trace sums, and the asymptotics of the associated trace formulas is unaffected.

chapter 28

In the intermittent case, the fixed point $\bar{0}$ does not provide any shadowing, and a statement such as

$$\Lambda_{1,0^{k+1}} \approx \Lambda_{1,0^k} \Lambda_{\bar{0}},$$

is meaningless. It seems therefore sensible to take out the factor $(1 - t_0) = 1 - z$ from the product representation of the dynamical zeta function (22.11), that is, to consider a pruned dynamical zeta function $1/\zeta_{inter}(z)$ defined by

$$1/\zeta(z) = (1 - z)1/\zeta_{inter}(z).$$

We saw in the last sections, that the zeta function $1/\zeta_{inter}(z)$ has all the nice properties we know from the hyperbolic case, that is, we can find a cycle expansion with - in the toy model case - vanishing curvature contributions and we can calculate dynamical properties like escape after having understood, how to handle the branch cut. But you might still be worried about leaving out the extra factor $1 - z$ all together. It turns out, that this is not only a matter of convenience, omitting the marginal $\bar{0}$ cycle is a dire necessity. The cycle weight $\Lambda_0^n = 1$ overestimates the corresponding interval length of $\mathcal{M}_{0^n}$ in the partition of the state space $\mathcal{M}$ by an increasing amount thus leading to wrong results when calculating escape. By leaving out the $\bar{0}$ cycle (and thus also the $\mathcal{M}_{0^n}$ contribution), we are guaranteed to get at least the right asymptotical behavior.

Note also, that if we are working with the spectral determinant (22.3), given in product form as

$$\det(1 - z\mathcal{L}) = \prod_p \prod_{m=0}^{\infty} \left(1 - \frac{z^{n_p}}{|\Lambda_p| \Lambda_p^m}\right),$$

for intermittent maps the marginal stable cycle has to be excluded. It introduces an (unphysical) essential singularity at $z = 1$ due the presence of a factor $(1 - z)^\infty$ stemming from the $\bar{0}$ cycle.

29.3 Intermittency for cyclists

Admittedly, the toy map is what is says - a toy model. The piece wise linearity of the map led to exact cancellations of the curvature contributions leaving only the fundamental terms. There are still infinitely many orbits included in the fundamental term, but the cycle weights were chosen in such a way that the zeta function could be written in closed form. For a smooth intermittent map this all will not be the case in general; still, we will argue that we have already seen almost all the fundamentally new features due to intermittency. What remains are technicalities - not necessarily easy to handle, but nothing very surprise any more.

In the following we will sketch, how to make cycle expansion techniques work for general 1-dimensional maps with a single isolated marginal fixed point. To keep the notation simple, we will consider two-branch maps with a complete binary symbolic dynamics as for example the Farey map, figure 29.3, or the repeller depicted in figure 29.2. We again assume that the behavior near the fixed point is given by (29.1). This implies that the stability of a family of periodic orbits approaching the marginally stable orbit, as for example the family 10^n , will increase only algebraically, that is we find again for large n

$$\frac{1}{\Lambda_{10^n}} \sim \frac{1}{n^{1+1/s}},$$

where s denotes the intermittency exponent.

When considering zeta functions or trace formulas, we again have to take out the marginal orbit $\bar{0}$; periodic orbit contributions of the form $t_{0^n 1}$ are now unbalanced and we arrive at a cycle expansion in terms of infinitely many fundamental

Table 29.1: Infinite alphabet versus the original binary alphabet for the shortest periodic orbit families. Repetitions of prime cycles ($11 = 1^2, 0101 = 01^2, \dots$) and their cyclic repeats ($110 = 101, 1110 = 1101, \dots$) are accounted for by cancelations and combination factors in the cycle expansion (29.30).

∞ – alphabet		binary alphabet				
		$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
1-cycles	n	1	10	100	1000	10000
2-cycles	mn					
	$1n$	11	110	1100	11000	110000
	$2n$	101	0101	10100	101000	1010000
	$3n$	1001	10010	100100	1001000	10010000
	$4n$	10001	100010	1000100	10001000	100010000
3-cycles	kmn					
	$11n$	111	1110	11100	111000	1110000
	$12n$	1101	11010	110100	1101000	11010000
	$13n$	11001	110010	1100100	11001000	110010000
	$21n$	1011	10110	101100	1011000	10110000
	$22n$	10101	101010	1010100	10101000	101010000
	$23n$	101001	1010010	10100100	101001000	1010010000
	$31n$	10011	100110	1001100	10011000	100110000
	$32n$	100101	1001010	10010100	100101000	1001010000
	$33n$	1001001	10010010	100100100	1001001000	10010010000

terms as for our toy map. This corresponds to moving from our binary symbolic dynamics to an infinite symbolic dynamics by making the identification

$$10^{n-1} \rightarrow n; \quad 10^{n-1}10^{m-1} \rightarrow nm; \quad 10^{n-1}10^{m-1}10^{k-1} \rightarrow nmk; \dots$$

see also table 29.1. The topological length of the orbit is thus no longer determined by the iterations of our two-branch map, but by the number of times the cycle goes from the right to the left branch. Equivalently, one may define a new map, for which all the iterations on the left branch are done in one step. Such a map is called an *induced map* and the topological length of orbits in the infinite alphabet corresponds to the iterations of this induced map.

exercise 15.1

For generic intermittent maps, curvature contributions in the cycle expanded zeta function will not vanish exactly. The most natural way to organize the cycle expansion is to collect orbits and pseudo orbits of the same topological length with respect to the infinite alphabet. Denoting cycle weights in the new alphabet as $t_{nm\dots} = t_{10^{n-1}10^{m-1}\dots}$, one obtains

$$\begin{aligned}
 \zeta^{-1} &= \prod_{p \neq 0} (1 - t_p) = 1 - \sum_{n=1}^{\infty} c_n \\
 &= 1 - \sum_{n=1}^{\infty} t_n - \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{2} (t_{mn} - t_m t_n) \\
 &\quad - \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \left(\frac{1}{3} t_{kmn} - \frac{1}{2} t_{km} t_n + \frac{1}{6} t_k t_m t_n \right) - \sum_{l=1}^{\infty} \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \dots
 \end{aligned} \tag{29.30}$$

The first sum is the fundamental term, which we have already seen in the toy

model, (29.10). The curvature terms c_n in the expansion are now e -fold infinite sums where the prefactors take care of double counting of prime periodic orbits.

Let us consider the fundamental term first. For generic intermittent maps, we can not expect to obtain an analytic expression for the infinite sum of the form

$$f(z) = \sum_{n=0}^{\infty} h_n z^n. \quad (29.31)$$

with algebraically decreasing coefficients

$$h_n \sim \frac{1}{n^\alpha} \quad \text{with} \quad \alpha > 0$$

To evaluate the sum, we face the same problem as for our toy map: the power series diverges for $z > 1$, that is, exactly in the ‘interesting’ region where poles, zeros or branch cuts of the zeta function are to be expected. By carefully subtracting the asymptotic behavior with the help of (29.11) or (29.12), one can in general construct an analytic continuation of $f(z)$ around $z = 1$ of the form

$$\begin{aligned} f(z) &\sim A(z) + (1-z)^{\alpha-1} B(z) & \alpha \notin \mathbb{N} \\ f(z) &\sim A(z) + (1-z)^{\alpha-1} \ln(1-z) & \alpha \in \mathbb{N}, \end{aligned} \quad (29.32)$$

where $A(z)$ and $B(z)$ are functions analytic in a disc around $z = 1$. We thus again find that the zeta function (29.30) has a branch cut along the real axis $\text{Re } z \geq 1$. From here on we can switch to auto-pilot and derive algebraic escape, decay of correlation and all the rest. We find in particular that the asymptotic behavior derived in (29.24) and (29.25) is a general result, that is, the survival probability is given asymptotically by

$$\Gamma_n \sim C \frac{1}{n^{1/s}} \quad (29.33)$$

for all 1-dimensional maps of the form (29.1). We have to work a bit harder if we want more detailed information like the prefactor C , exponential precursors given by zeros or poles of the dynamical zeta function or higher order corrections. This information is buried in the functions $A(z)$ and $B(z)$ or more generally in the analytically continued zeta function. To get this analytic continuation, one may follow either of the two different strategies which we will sketch next.

29.3.1 Resummation

One way to get information about the zeta function near the branch cut is to derive the leading coefficients in the Taylor series of the functions $A(z)$ and $B(z)$ in (29.32) at $z = 1$. This can be done in principle, if the coefficients h_n in sums like

(29.31) are known (as for our toy model). One then considers a resummation of the form

$$\sum_{j=0}^{\infty} h_j z^j = \sum_{j=0}^{\infty} a_j (1-z)^j + (1-z)^{\alpha-1} \sum_{j=0}^{\infty} b_j (1-z)^j, \quad (29.34)$$

and the coefficients a_j and b_j are obtained in terms of the h_j 's by expanding $(1-z)^j$ and $(1-z)^{j+\alpha-1}$ on the right hand side around $z = 0$ using (29.11) and equating the coefficients.

In practical calculations one often has only a finite number of coefficients h_j , $0 \leq j \leq N$, which may have been obtained by finding periodic orbits and their stabilities numerically. One can still design a resummation scheme for the computation of the coefficients a_j and b_j in (29.34). We replace the infinite sums in (29.34) by finite sums of increasing degrees n_a and n_b , and require that

$$\sum_{i=0}^{n_a} a_i (1-z)^i + (1-z)^{\alpha-1} \sum_{i=0}^{n_b} b_i (1-z)^i = \sum_{i=0}^N h_i z^i + O(z^{N+1}). \quad (29.35)$$

One proceeds again by expanding the right hand side around $z = 0$, skipping all powers z^{N+1} and higher, and then equating coefficients. It is natural to require that $|n_b + \alpha - 1 - n_a| < 1$, so that the maximal powers of the two sums in (29.35) are adjacent. If one chooses $n_a + n_b + 2 = N + 1$, then, for each cutoff length N , the integers n_a and n_b are uniquely determined from a linear system of equations. The price we pay is that the so obtained coefficients depend on the cutoff N . One can now study convergence of the coefficients a_j , and b_j , with respect to increasing values of N , or various quantities derived from a_j and b_j . Note that the leading coefficients a_0 and b_0 determine the prefactor C in (29.33), cf. (29.23). The resummed expression can also be used to compute zeros, inside or outside the radius of convergence of the cycle expansion $\sum h_j z^j$.

The scheme outlined in this section tacitly assumes that a representation of form (29.32) holds in a disc of radius 1 around $z = 1$. Convergence is improved further if additional information about the asymptotics of sums like (29.31) is used to improve the ansatz (29.34).

29.3.2 Analytical continuation by integral transformations

We will now introduce a method which provides an analytic continuation of sums of the form (29.31) without explicitly relying on an ansatz (29.34). The main idea is to rewrite the sum (29.31) as a sum over integrals with the help of the Poisson summation formula and find an analytic continuation of each integral by contour deformation. In order to do so, we need to know the n dependence of the coefficients $h_n \equiv h(n)$ explicitly for all n . If the coefficients are not known analytically, one may proceed by approximating the large n behavior in the form

$$h(n) = n^{-\alpha} (C_1 + C_2 n^{-1} + \dots), \quad n \neq 0,$$

and determine the constants C_i numerically from periodic orbit data. By using the Poisson resummation identity

$$\sum_{n=-\infty}^{\infty} \delta(x - n) = \sum_{m=-\infty}^{\infty} \exp(2\pi i m x), \quad (29.36)$$

we may write the sum as (29.31)

$$f(z) = \frac{1}{2}h(0) + \sum_{m=-\infty}^{\infty} \int_0^{\infty} dx e^{2\pi i m x} h(x) z^x. \quad (29.37)$$

The continuous variable x corresponds to the discrete summation index n and it is convenient to write $z = r \exp(i\sigma)$ from now on. The integrals are still not convergent for $r > 0$, but an analytical continuation can be found by considering the contour integral, where the contour goes out along the real axis, makes a quarter circle to either the positive or negative imaginary axis and goes back to zero. By letting the radius of the circle go to infinity, we essentially rotate the line of integration from the real onto the imaginary axis. For the $m = 0$ term in (29.37), we transform $x \rightarrow ix$ and the integral takes on the form

$$\int_0^{\infty} dx h(x) r^x e^{ix\sigma} = i \int_0^{\infty} dx h(ix) r^{ix} e^{-x\sigma}.$$

The integrand is now exponentially decreasing for all $r > 0$ and $\sigma \neq 0$ or 2π . The last condition reminds us again of the existence of a branch cut at $\text{Re } z \geq 1$. By the same technique, we find the analytic continuation for all the other integrals in (29.37). The real axis is then rotated according to $x \rightarrow \text{sign}(m)ix$ where $\text{sign}(m)$ refers to the sign of m .

$$\int_0^{\infty} dx e^{\pm 2\pi i |m|x} h(x) r^x e^{ix\sigma} = \pm i \int_0^{\infty} dx h(\pm ix) r^{\pm ix} e^{-x(2\pi |m| \pm \sigma)}.$$

Changing summation and integration, we can carry out the sum over $|m|$ explicitly and one finally obtains the compact expression

$$\begin{aligned} f(z) &= \frac{1}{2}h(0) + i \int_0^{\infty} dx h(ix) r^{ix} e^{-x\sigma} \\ &+ i \int_0^{\infty} dx \frac{e^{-2\pi x}}{1 - e^{-2\pi x}} \left[h(ix) r^{ix} e^{-x\sigma} - h(-ix) r^{-ix} e^{x\sigma} \right]. \end{aligned} \quad (29.38)$$

The transformation from the original sum to the two integrals in (29.38) is exact for $r \leq 1$, and provides an analytic continuation for $r > 0$. The expression (29.38) is especially useful for an efficient numerical calculations of a dynamical zeta function for $|z| > 1$, which is essential when searching for its zeros and poles.

29.3.3 Curvature contributions

So far, we have discussed only the fundamental term $\sum_{n=1}^{\infty} t_n$ in (29.30), and showed how to deal with such power series with algebraically decreasing coefficients. The fundamental term determines the main structure of the zeta function

in terms of the leading order branch cut. Corrections to both the zeros and poles of the dynamical zeta function as well as the leading and subleading order terms in expansions like (29.32) are contained in the curvature terms in (29.30). The first curvature correction is the 2-cycle sum

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{2} (t_{mn} - t_m t_n),$$

with algebraically decaying coefficients which again diverge for $|z| > 1$. The analytically continued curvature terms have as usual branch cuts along the positive real z axis. Our ability to calculate the higher order curvature terms depends on how much we know about the cycle weights t_{mn} . The form of the cycle stability (29.6) suggests that t_{mn} decrease asymptotically as

$$t_{mn} \sim \frac{1}{(nm)^{1+1/s}} \quad (29.39)$$

for 2-cycles, and in general for n -cycles as

$$t_{m_1 m_2 \dots m_n} \sim \frac{1}{(m_1 m_2 \dots m_n)^{1+1/s}}.$$

If we happen to know the cycle weights $t_{m_1 m_2 \dots m_n}$ analytically, we may proceed as in sect. 29.3.2, transform the multiple sums into multiple integrals and rotate the integration contours.

We have reached the edge of what has been accomplished so far in computing and what is worth the dynamical zeta functions from periodic orbit data. In the next section, we describe a probabilistic method applicable to intermittent maps which does not rely on periodic orbits.

29.3.4 Stability ordering for intermittent flows

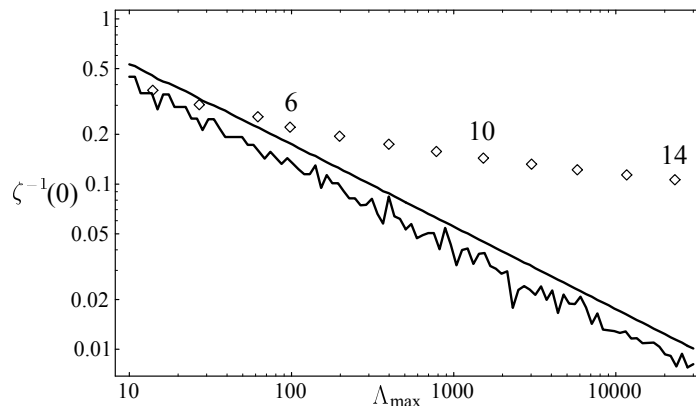


Longer but less unstable cycles can give larger contributions to a cycle expansion than short but highly unstable cycles. In such situations, truncation by length may require an exponentially large number of very unstable cycles before a significant longer cycle is first included in the expansion. This situation is best illustrated by intermittent maps. The simplest of these is the Farey map

$$f(x) = \begin{cases} f_0 = x/(1-x) & 0 \leq x \leq 1/2 \\ f_1 = (1-x)/x & 1/2 \leq x \leq 1 \end{cases}, \quad (29.40)$$

For the Farey map, the symbolic dynamics is of complete binary type, so the lack of shadowing is not due to the lack of a finite grammar, but rather to the intermittency caused by the existence of the marginal fixed point $x_0 = 0$, for which the stability multiplier is $\Lambda_0 = 1$. This fixed point does not participate directly in the dynamics and is omitted from cycle expansions. Its presence is, however, very much felt instead in the stabilities of neighboring cycles with n consecutive iterates of the symbol 0, whose stability falls off only as $\Lambda \sim n^2$, in contrast to

Figure 29.8: Comparison of cycle expansion truncation schemes for the Farey map (29.40); the deviation of the truncated cycles expansion for $|1/\zeta_N(0)|$ from the exact flow conservation value $1/\zeta(0) = 0$ is a measure of the accuracy of the truncation. The jagged line is the logarithm of the stability ordering truncation error; the smooth line is smoothed according to sect. 23.7.2; the diamonds indicate the error due to the topological period truncation, with the maximal cycle period N shown. They are placed along the stability cutoff axis at points determined by the condition that the total number of cycles is the same for both truncation schemes.



the most unstable cycles with n consecutive 1's, which are exponentially unstable, $|\Lambda_{01^n}| \sim [(\sqrt{5} + 1)/2]^{2n}$.

The symbolic dynamics is of complete binary type. A quick count in the style of sect. 18.7.2 leads to a total of 74,248,450 prime cycles of period 30 or less, not including the marginal point $x_0 = 0$. Evaluating a cycle expansion to this order is an impressive computational feat. However, stability of the least unstable cycle omitted is roughly $\Lambda_{10^{30}} \sim 30^2 = 900$, so it yields a 0.1% correction. The situation may be much worse than this estimate suggests, because the next 10^{31} cycle contributes a similar amount, and could easily reinforce the error. Adding up all such omitted terms, we arrive at an estimated error of about 3%, for a cycle-period truncated cycle expansion based on more than 10^9 pseudo-cycle terms! On the other hand, if one truncates by stability at $\Lambda_{\max} = 3000$, only 409 prime cycles suffice to attain the same accuracy of about a 3% error, figure 29.8.

As the Farey map maps the unit interval onto itself, the leading eigenvalue of the Perron-Frobenius operator should equal $s_0 = 0$, so $1/\zeta(0) = 0$. Deviations from this exact result give an indication of the convergence of a given cycle expansion. Errors corresponding to different truncation schemes are indicated in figure 29.8. We see that topological period truncation schemes are hopelessly bad in this case; stability length truncations are somewhat better, but still rather bad. In simple cases like this one, where intermittency is caused by a single marginal fixed point, convergence can be improved by going to infinite alphabets.

29.4 BER zeta functions



So far we have focused on 1-d models as the simplest setting in which to investigate dynamical implications of marginal fixed points. We now take an altogether different track and describe how probabilistic methods may be employed in order to write down approximate dynamical zeta functions for intermittent systems.

We will discuss the method in a very general setting, for a flow in arbitrary

dimension. The key idea is to introduce a surface of section $\mathcal{P}$ such that all trajectories traversing this section will have spent some time both near the marginal stable fixed point and in the chaotic phase. An important quantity in what follows is (3.7), the *first return time* $\tau(x)$, or the time of flight of a trajectory starting in x to the next return to the surface of section $\mathcal{P}$. The period of a periodic orbit p intersecting the $\mathcal{P}$ section n_p times is

$$T_p = \sum_{k=0}^{n_p-1} \tau(f^k(x_p)),$$

where $f(x)$ is the Poincaré map, and $x_p \in \mathcal{P}$ is a periodic point. The dynamical zeta function (22.11)

$$1/\zeta(z, s, \beta) = \prod_p \left(1 - \frac{z^{n_p} e^{\beta A_p - s T_p}}{|\Lambda_p|} \right), \quad A_p = \sum_{k=0}^{n_p-1} a(f^k(x_p)), \quad (29.41)$$

associated with the observable $a(x)$ captures the dynamics of both the flow *and* the Poincaré map. The dynamical zeta function for the flow is obtained as $1/\zeta(s, \beta) = 1/\zeta(1, s, \beta)$, and the dynamical zeta function for the discrete time Poincaré map is $1/\zeta(z, \beta) = 1/\zeta(z, 0, \beta)$.

chapter 20

Our basic assumption will be *probabilistic*. We assume that the chaotic interludes render the consecutive *return* (or *recurrence*) *times* $T(x_i)$, $T(x_{i+1})$ and observables $a(x_i)$, $a(x_{i+1})$ effectively uncorrelated. Consider the quantity $e^{\beta A(x_0, n) - s T(x_0, n)}$ averaged over the surface of section $\mathcal{P}$. With the above probabilistic assumption the large n behavior is

$$\langle e^{\beta A(x_0, n) - s T(x_0, n)} \rangle_{\mathcal{P}} \sim \left(\int_{\mathcal{P}} e^{\beta a(x) - s \tau(x)} \rho(x) dx \right)^n,$$

where $\rho(x)$ is the invariant density of the Poincaré map. This type of behavior is equivalent to there being only one zero $z_0(s, \beta) = \int e^{\beta a(x) - s \tau(x)} \rho(x) dx$ of $1/\zeta(z, s, \beta)$ in the z - β plane. In the language of Ruelle-Pollicott resonances this means that there is an infinite gap to the first resonance. This in turn implies that $1/\zeta(z, s, \beta)$ may be written as

remark 20.2

$$1/\zeta(z, s, \beta) = z - \int_{\mathcal{P}} e^{\beta a(x) - s \tau(x)} \rho(x) dx,$$

where we have neglected a possible analytic and non-zero prefactor. The dynamical zeta function of the flow is now

$$1/\zeta(s, \beta) = 1/\zeta(1, s, \beta) = 1 - \int_{\mathcal{P}} e^{\beta a(x)} \rho(x) e^{-s \tau(x)} dx. \quad (29.42)$$

Normally, the best one can hope for is a finite gap to the leading resonance of the Poincaré map. with the above dynamical zeta function only approximatively valid. As it is derived from an approximation due to Baladi, Eckmann, and Ruelle, we shall refer to it as the BER zeta function $1/\zeta_{\text{BER}}(s, \beta)$ in what follows.

A central role is played by the probability distribution of return times

$$\psi(\tau) = \int_{\mathcal{P}} \delta(\tau - \tau(x)) \rho(x) dx$$

The BER zeta function at $\beta = 0$ is then given in terms of the Laplace transform of this distribution

exercise 24.6

$$1/\zeta_{\text{BER}}(s) = 1 - \int_0^\infty \psi(\tau) e^{-s\tau} d\tau.$$

exercise 29.5



example 29.1
p. 606



example 29.2
p. 606

It may seem surprising that the BER approximation produces exact results in the two examples above. The reason for this peculiarity is that both these systems are piecewise linear and have complete Markov partitions. As long as the map is piecewise linear and complete, and the probabilistic approximation is exactly fulfilled, the cycle expansion curvature terms vanish. The BER zeta function and the fundamental part of a cycle expansion discussed in sect. 23.1.1 are indeed intricately related, but not identical in general. In particular, note that the BER zeta function obeys the flow conservation sum rule (23.17) by construction, whereas the fundamental part of a cycle expansion as a rule does not.

Résumé

The presence of marginally stable fixed points and cycles changes the analytic structure of dynamical zeta functions and the rules for constructing cycle expansions. The marginal orbits have to be omitted, and the cycle expansions now need to include families of infinitely many longer and longer unstable orbits which accumulate toward the marginally stable cycles. Correlations for such nonhyperbolic systems may decay algebraically with the decay rates controlled by the branch cuts of dynamical zeta functions. Compared to pure hyperbolic systems, the physical consequences are drastic: exponential decays are replaced by slow power-law decays, and transport properties, such as the diffusion may become anomalous.

Commentary

Remark 29.1. What about the evolution operator formalism? The main virtue of evolution operators was their semigroup property (20.16). This was natural for hyperbolic systems where instabilities grow exponentially, and evolution operators capture this behavior due to their multiplicative nature. Whether the evolution operator formalism is a good way to capture the slow, power law instabilities of intermittent dynamics is less clear. The approach taken here leads us to a formulation in terms of *dynamical zeta functions* rather than spectral determinants, circumventing evolution operators altogether. It is not known if the spectral determinants formulation would yield any benefits when applied to intermittent chaos. Some results on spectral determinants and intermittency can be

found in ref. [24]. A useful mathematical technique to deal with isolated marginally stable fixed point is that of *inducing*, that is, replacing the intermittent map by a completely hyperbolic map with infinite alphabet and redefining the discrete time; we have used this method implicitly by changing from a finite to an infinite alphabet. We refer to refs. [13, 21, 22] for detailed discussions of this technique, as well as applications to 1-dimensional maps.

Remark 29.2. Intermittency. Intermittency was discovered by Manneville and Pomeau [15] in their study of the Lorenz system. They demonstrated that in neighborhood of parameter value $r_c = 166.07$ the mean duration of the periodic motion scales as $(r - r_c)^{1/2}$. In ref. [20] they explained this phenomenon in terms of a 1-dimensional map (such as (29.1)) near tangent bifurcation, and classified possible types of intermittency.

Piecewise linear models like the one considered here have been studied by Gaspard and Wang [9, 26, 27]. The escape problem has here been treated following ref. [6], resummations following ref. [5]. The proof of the bound (29.27) is given in P. Dahlqvist's notes, see ChaosBook.org/extras/PDahlqvistEscape.pdf.

Farey map (29.40) has been studied widely in the context of intermittent dynamics, for example in refs. [1, 7, 16–19, 24]. The Fredholm determinant and the dynamical zeta functions for the Farey map (29.40) and the related Gauss shift map (19.45) have been studied by Mayer [17]. He relates the continued fraction transformation to the Riemann zeta function, and constructs a Hilbert space on which the evolution operator is self-adjoint, and its eigenvalues are exponentially spaced, just as for the dynamical zeta functions [23] for “Axiom A” hyperbolic systems.

Remark 29.3. Tauberian theorems. In this chapter we used Tauberian theorems for power series and Laplace transforms: Feller's monograph [8] is a highly recommended introduction to these methods.

Remark 29.4. Probabilistic methods, BER zeta functions. Probabilistic description of intermittent chaos was introduced by Geisal and Thomae [11]. The BER approximation studied here is inspired by Baladi, Eckmann and Ruelle [2], with further developments in refs. [3, 4].

References

- [1] R. Artuso, E. Aurell, and P. Cvitanović, “Recycling of strange sets: II. Applications”, *Nonlinearity* **3**, 361–386 (1990).
- [2] V. Baladi, J.-P. Eckmann, and D. Ruelle, “Resonances for intermittent systems”, *Nonlinearity* **2**, 119–135 (1989).
- [3] P. Dahlqvist, “Determination of resonance spectra for bound chaotic systems”, *J. Phys. A* **27**, 763–785 (1994).
- [4] P. Dahlqvist, “Approximate zeta functions for the Sinai billiard and related systems”, *Nonlinearity* **8**, 11 (1995).
- [5] P. Dahlqvist, “Computing the topological pressure for intermittent maps”, *J. Phys. A* **30**, L351–L358 (1997).

- [6] P. Dahlqvist, “Escape from intermittent repellers: Periodic orbit theory for crossover from exponential to algebraic decay”, *Phys. Rev. E* **60**, 6639–6644 (1999).
- [7] C. P. Dettmann and P. Cvitanović, “Cycle expansions for intermittent diffusion”, *Phys. Rev. E* **56**, 6687–6692 (1997).
- [8] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. (Wiley, 1971).
- [9] P. Gaspard and X.-J. Wang, “Sporadicity: Between periodic and chaotic dynamical behaviors”, *Proc. Natl. Acad. Sci. USA* **85**, 4591–4595 (1988).
- [10] T. Geisel and J. Nierwetberg, “Onset of diffusion and universal scaling in chaotic systems”, *Phys. Rev. Lett.* **48**, 7 (1982).
- [11] T. Geisel and S. Thomaes, “Anomalous diffusion in intermittent chaotic systems”, *Phys. Rev. Lett.* **52**, 1936 (1984).
- [12] S. Grossmann and H. Fujisaka, “Diffusion in discrete nonlinear dynamical systems”, *Phys. Rev. A* **26**, 1779–1782 (1982).
- [13] S. Isola, “Renewal sequences and intermittency”, *J. Stat. Phys.* **97**, 263–280 (1999).
- [14] R. Lombardi, Laurea thesis, MA thesis (Università degli studi di Milano, 1993).
- [15] P. Manneville and Y. Pomeau, “Intermittency and the Lorenz model”, *Phys. Lett. A* **75**, 1–2 (1979).
- [16] D. Mayer and G. Roepstorff, “On the relaxation time of Gauss’s continued-fraction map I. The Hilbert space approach (Koopmanism)”, *J. Stat. Phys.* **47**, 149–171 (1987).
- [17] D. H. Mayer, “On a ζ function related to the continued fraction transformation”, *Bull. Soc. Math. Fr.* **104**, 195–203 (1976).
- [18] D. H. Mayer, *The Ruelle-Araki Transfer Operator in Classical Statistical Mechanics* (Springer, Berlin, 1980).
- [19] D. H. Mayer, “Continued fractions and related transformations”, in *Ergodic Theory, Symbolic Dynamics, and Hyperbolic Spaces*, edited by T. Bedford, M. S. Keane, and C. Series (Oxford Univ. Press, Oxford, 1991).
- [20] Y. Pomeau and P. Manneville, “Intermittent transition to turbulence in dissipative dynamical systems”, *Commun. Math. Phys.* **74**, 189 (1980).
- [21] T. Prellberg, Maps of the Interval with Indifferent Fixed Points: Thermodynamic Formalism and Phase Transitions, PhD thesis (Virginia Polytechnic Inst., 1991).
- [22] T. Prellberg and J. Slawny, “Maps of intervals with indifferent fixed points: Thermodynamic formalism and phase transitions”, *J. Stat. Phys.* **66**, 503–514 (1992).
- [23] D. Ruelle, “Zeta-functions for expanding maps and Anosov flows”, *Inv. Math.* **34**, 231–242 (1976).
- [24] H. H. Rugh, “Intermittency and regularized Fredholm determinants”, *Inv. Math.* **135**, 1–24 (1999).

- [25] M. Schell, S. Fraser, and R. Kapral, “Diffusive dynamics in systems with translational symmetry: A one-dimensional-map model”, *Phys. Rev. A* **26**, 504–521 (1982).
- [26] X.-J. Wang, “Abnormal fluctuations and thermodynamic phase transitions in dynamical systems”, *Phys. Rev. A* **39**, 3214–3217 (1989).
- [27] X.-J. Wang, “Statistical physics of temporal intermittency”, *Phys. Rev. A* **40**, 6647–6661 (1989).

29.5 Examples

Example 29.1. Return times for the Bernoulli map. For the Bernoulli shift map (28.19)

$$x \mapsto f(x) = 2x \bmod 1,$$

one easily derives the distribution of return times

$$\psi_n = \frac{1}{2^n} \quad n \geq 1.$$

The BER zeta function becomes (by the discrete Laplace transform (21.8))

$$\begin{aligned} 1/\zeta_{\text{BER}}(z) &= 1 - \sum_{n=1}^{\infty} \psi_n z^n = 1 - \sum_{n=1}^{\infty} \frac{z^n}{2^n} \\ &= \frac{1-z}{1-z/2} = \zeta^{-1}(z)/(1-z/\Lambda_0). \end{aligned} \quad (29.43)$$

Thanks to the uniformity of the piecewise linear map measure (19.40) the “approximate” zeta function is in this case the *exact* dynamical zeta function, with the periodic point $\bar{0}$ pruned. [return to p. 602](#)

Example 29.2. Return times for the model of sect. 29.2.1. For the toy model of sect. 29.2.1 one gets $\psi_1 = |\mathcal{M}_1|$, and $\psi_n = |\mathcal{M}_n|(1-b)/(1-a)$, for $n \geq 2$, leading to a BER zeta function

$$1/\zeta_{\text{BER}}(z) = 1 - z|\mathcal{M}_1| - \sum_{n=2}^{\infty} |\mathcal{M}_n|z^n,$$

which again coincides with the exact result, (29.10). [return to p. 602](#)

29.6 Exercises

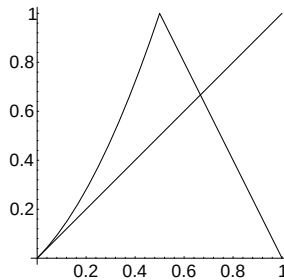
29.1. **Integral representation of Jonquière functions.** Check the integral representation

$$J(z, \alpha) = \frac{z}{\Gamma(\alpha)} \int_0^\infty d\xi \frac{\xi^{\alpha-1}}{e^\xi - z} \quad \text{for } \alpha > 0. \quad (29.44)$$

Note how the denominator is connected to Bose-Einstein distribution. Compute $J(x + i\epsilon) - J(x - i\epsilon)$ for a real $x > 1$.

29.2. **Power law correction to a power law.** Expand (29.20) further and derive the leading power law correction to (29.23).

29.3. **Power-law fall off.** In cycle expansions the stabilities of orbits do not always behave in a geometric fashion. Consider the map f



This map behaves as $f \rightarrow x$ as $x \rightarrow 0$. Define a symbolic dynamics for this map by assigning 0 to the points that land on the interval $[0, 1/2)$ and 1 to the points that land on $(1/2, 1]$. Show that the stability of orbits that spend a long time on the 0 side goes as n^2 . In particular, show that

$$\Lambda_{\underbrace{00\dots0}_n 1} \sim n^2$$

29.4. **Power law fall-off of Floquet multipliers in the stadium billiard.**  From the cycle expansions point of view, the most important consequence of the shear in $\mathbf{J}^n$ for long sequences of rotation bounces n_k in (9.14) is that the Λ_n grows only as a power law in number of bounces:

$$\Lambda_n \propto n_k^2. \quad (29.45)$$

Check.

29.5. **Probabilistic zeta function for maps.** Derive the probabilistic zeta function for a map with recurrence distribution ψ_n .

29.6. **Accelerated diffusion.** Consider a map h , such that $\hat{h} = \hat{f}$, but now running branches are turned into standing branches and vice versa, so that 1, 2, 3, 4 are standing while 0 leads to both positive and negative jumps. Build the corresponding dynamical zeta function and show that

$$\sigma^2(t) \sim \begin{cases} t & \text{for } \alpha > 2 \\ t \ln t & \text{for } \alpha = 2 \\ t^{3-\alpha} & \text{for } \alpha \in (1, 2) \\ t^2 / \ln t & \text{for } \alpha = 1 \\ t^2 & \text{for } \alpha \in (0, 1) \end{cases}$$

- 29.7. **Anomalous diffusion (hyperbolic maps).** Anomalous diffusive properties are associated to deviations from linearity of the variance of the phase variable we are looking at: this means the diffusion constant (20.30) either vanishes or diverges. We briefly illustrate in this exercise how the local local properties of a map are crucial to account for anomalous behavior even for hyperbolic systems.

Consider a class of piecewise linear maps, relevant to the problem of the onset of diffusion, defined by

$$f_\epsilon(x) = \begin{cases} \Lambda x & \text{for } x \in [0, x_1^+] \\ a - \Lambda_{\epsilon,\gamma}|x - x^+| & \text{for } x \in [x_1^+, x_2^+] \\ 1 - \Lambda'(x - x_2^+) & \text{for } x \in [x_2^+, x_1^-] \\ 1 - a + \Lambda_{\epsilon,\gamma}|x - x^-| & \text{for } x \in [x_1^-, x_2^-] \\ 1 + \Lambda(x - 1) & \text{for } x \in [x_2^-, 1] \end{cases}$$

where $\Lambda = (1/3 - \epsilon^{1/\gamma})^{-1}$, $\Lambda' = (1/3 - 2\epsilon^{1/\gamma})$, $\Lambda_{\epsilon,\gamma} = \epsilon^{1-1/\gamma}$, $a = 1 + \epsilon$, $x^+ = 1/3$, $x_1^+ = x^+ - \epsilon^{1/\gamma}$, $x_2^+ = x^+ + \epsilon^{1/\gamma}$, and the usual symmetry properties (24.23) are satisfied.

Thus this class of maps is characterized by two escaping windows (through which the diffusion process may take place) of size $2\epsilon^{1/\gamma}$: the exponent γ mimicks the order of the maximum for a continuous map, while piecewise linearity, besides making curvatures vanish and leading to finite cycle expansions, prevents the appearance of stable cycles. The symbolic dynamics is easily described once we consider a sequence of parameter values $\{\epsilon_m\}$, where $\epsilon_m = \Lambda^{-(m+1)}$: we then partition the unit interval though the sequence of points $0, x_1^+, x^+, x_2^+, x_1^-, x^-, x_2^-, 1$ and label the corresponding sub-intervals 1, $s_a, s_b, 2, d_b, d_a, 3$: symbolic dynamics is described by an unrestricted grammar over the following set of symbols

$$\{1, 2, 3, s_\# \cdot 1^i, d_\# \cdot 3^k\} \quad \# = a, b \quad i, k = m, m+1, m+2, \dots$$

This leads to the following dynamical zeta function:

$$\zeta_0^{-1}(z, \alpha) = 1 - \frac{2z}{\Lambda} - \frac{z}{\Lambda'} - 4 \cosh(\alpha) \epsilon_m^{1/\gamma-1} \frac{z^{m+1}}{\Lambda^m} \left(1 - \frac{z}{\Lambda}\right)^{-1}$$

from which, by (24.9) we get

$$D = \frac{2\epsilon_m^{1/\gamma-1} \Lambda^{-m} (1 - 1/\Lambda)^{-1}}{1 - \frac{2}{\Lambda} - \frac{1}{\Lambda'} - 4\epsilon_m^{1/\gamma-1} \left(\frac{m+1}{\Lambda^m (1-1/\Lambda)} + \frac{1}{\Lambda^{m+1} (1-1/\Lambda)^2} \right)} \quad (29.46)$$

The main interest in this expression is that it allows exploring how D vanishes in the $\epsilon \mapsto 0$ ($m \mapsto \infty$) limit: as a matter of fact, from (29.46) we get the asymptotic behavior $D \sim \epsilon^{1/\gamma}$, which shows how the onset of diffusion is governed by the order of the map at its maximum.

Remark 29.5. Onset of diffusion for continuous maps. The zoology of behavior for continuous maps at the onset of diffusion is described in refs. [10, 12, 25]: our treatment for piecewise linear maps was introduced in ref. [14].

Chapter 30

Turbulence?

I am an old man now, and when I die and go to Heaven there are two matters on which I hope enlightenment. One is quantum electro-dynamics and the other is turbulence of fluids. About the former, I am rather optimistic.

—Sir Horace Lamb

THERE IS ONLY ONE honorable cause that would justify sweating through so much formalism - this is but the sharpening of a pencil in order that we may attack the Navier-Stokes equation,

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{\nabla p}{\rho} + \nu \nabla^2 \mathbf{u} + \mathbf{f}, \quad (30.1)$$

and solve the problem of turbulence.

Being realistic, we are not so foolhardy to immediately plunge into *the* problem – there are too many dimensions and indices. Instead, we start small, in one spatial dimension, $\mathbf{u} \rightarrow u$, $\mathbf{u} \cdot \nabla \mathbf{u} \rightarrow u \partial_x u$, assume constant density ρ , forget about the pressure p , and so on. This line of reasoning, as well as many other equally sensible threads of thought, such as the amplitude equations obtained via weakly nonlinear stability analysis of steady flows, leads to a small set of frequently studied nonlinear PDE models, like the one that we turn to now. You only need chapters 2 to 5 and chapters 14 to 15 to get started.



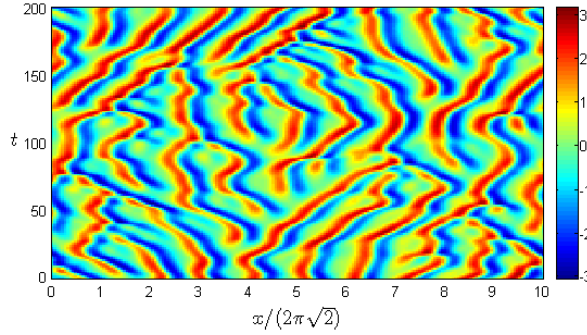
30.1 Configuration space: a fluttering flame front

Romeo: ‘Misshapen chaos of well seeming forms!’

—W. Shakespeare, *Romeo and Juliet*, Act I, Scene I

The Kuramoto-Sivashinsky equation is one of the simplest nonlinear systems used to describe ‘turbulence’ (often modestly referred to here as ‘spatiotemporally chaotic behavior’). This equation was first derived to model a laminar flame

Figure 30.1: Spacetime [heat map](#) of a typical “turbulent” solution of the Kuramoto-Sivashinsky equation, system size $\tilde{L} = 10\sqrt{2} \approx 88.86$. The color indicates the value of u at any given position and time. The x coordinate is plotted in units of the most unstable wavelength $2\pi\sqrt{2}$, which is empirically approximately the mean wavelength of the turbulent flow. The dynamics is typical of a large system, in this case approximately 10 mean wavelengths wide (from ref. [11]).



front, such as the flickering of a gas flame on your kitchen stove, figure 2.7. The time evolution of the ‘flame front velocity’ $u = u(x, t)$ on a periodic domain $u(x, t) = u(x + L, t)$ is given by

$$u_t + \frac{1}{2}(u^2)_x + u_{xx} + \nu u_{xxxx} = 0, \quad x \in [0, L]. \quad (30.2)$$

In this equation $t \geq 0$ is the time and x is the spatial coordinate. The subscripts x and t denote partial derivatives with respect to x and t : $u_t = \partial u / \partial t$, u_{xxxx} stands for the 4th spatial derivative of $u = u(x, t)$ at position x and time t . In what follows we use interchangeably the ‘dimensionless system size’ $\tilde{L}$, or the periodic domain size $L = 2\pi\tilde{L}$, as the system parameter. We take note, as in the Navier-Stokes equation (30.1), of the ‘inertial’ term $u\partial_x u$, the ‘anti-diffusive’ term $\partial_x^2 u$ (with a “wrong” sign), ‘(hyper-)viscosity’ ν , etc..

In what follows we will analyze -step by step- PDEs, using methods developed above for finite dimensional dynamical systems, a discussion started in sect. 2.4 *Life in extreme dimensions*. First we discuss PDEs as fields defined over configuration space. Once we go over to the state space description the techniques developed for analysis of ODEs will go over to PDEs, as is.

30.1.1 Symmetries of Kuramoto-Sivashinsky equation

The Kuramoto-Sivashinsky equation (30.2) is space translationally invariant, time translationally invariant, and invariant under reflection $x \rightarrow -x$, $u \rightarrow -u$. Comparing u_t and $(u^2)_x$ terms we note that u has dimensions of $[x]/[t]$, hence u is the ‘velocity’, rather than the ‘height’ of the flame front. Indeed, the Kuramoto-Sivashinsky equation is Galilean invariant: if $u(x, t)$ is a solution, then $v + u(x - vt, t)$, with v an arbitrary constant velocity, is also a solution. Without loss of generality, in our calculations we shall work in the zero mean velocity frame

$$\int dx u = 0. \quad (30.3)$$

In terms of the system size L , the only length scale available, the dimensions of terms in (30.2) are $[x] = L$, $[t] = L^2$, $[u] = L^{-1}$, $[\nu] = L^2$. Scaling out the “viscosity” parameter ν by $x \rightarrow x\nu^{1/2}$, $t \rightarrow t\nu$, $u \rightarrow u\nu^{-1/2}$, brings the Kuramoto-Sivashinsky equation (30.2) to a non-dimensional form (for the “integral”

form of the equations, see remark 30.3)

$$u_t + u u_x + u_{xx} + u_{xxx} = 0, \quad x \in [0, L\nu^{-1/2}) = [0, 2\pi\tilde{L}). \quad (30.4)$$

In this way we trade in both the “viscosity” ν and the system size L for a single dimensionless parameter

$$\tilde{L} = L/(2\pi\sqrt{\nu}) \quad (30.5)$$

which plays the role of a “Reynolds number” for the Kuramoto-Sivashinsky system. Some authors use L as the system parameter, with ν fixed to 1, and others vary ν with L fixed to either 1 or 2π . Physically, varying L is the right thing to do if one is interested in taking L large, and studying ‘spatio-temporal chaos.’ In what follows we state results of all calculations in units of the dimensionless parameter $\tilde{L}$.

The 2π factor in (30.5) is motivated by the stability analysis of the quiescent $u(x, t) = 0$ equilibrium. First, in these units the $u(x, t) = 0$ equilibrium bifurcates at each integer value $\tilde{L} = n$, see figure 30.2. Second, in these units the $2\pi\sqrt{2}$ mean wavelength estimate, confirmed by simulations such as figure 30.1, follows from the analytic formula for the stability exponents (5.2) of the $u(x, t) = 0$ stationary solution, see the Fourier modes form of the Kuramoto-Sivashinsky equation (30.14),

$$\lambda^{(k)} = q_k^2 - q_k^4, \quad q_k = k/\tilde{L}. \quad (30.6)$$

The most unstable mode, nearest to the maximum of this stability formula $k = \tilde{L}/\sqrt{2}$, sets the mean wavelength $\sqrt{2}$ of the KS ‘turbulent’ dynamics, see figure 30.8 and example 30.5.



example 30.5
p. 633

The time units also have to be rescaled; for example, if T_p^* is a period of a periodic solution of (30.2) with a given ν and $L = 2\pi$, then the corresponding solution of the non-dimensionalized (30.4) has period

$$T_p = T_p^*/\nu. \quad (30.7)$$

The term $(u^2)_x$ in (30.2) makes this a *nonlinear system*. This is one of the simplest conceivable nonlinear PDE, playing the role in the theory of spatially extended systems a bit like the role that the x^2 nonlinearity plays in the dynamics of iterated mappings. The time evolution of a typical solution of the Kuramoto-Sivashinsky system is illustrated by figure 30.1.

example 3.7
section 3.3
remark 30.1

Let $\mathbb{U}$ be the space of real-valued velocity fields periodic and square integrable on the interval $\Omega = [-L/2, L/2]$,

$$\mathbb{U} = \{u \in L^2(\Omega) \mid u(x) = u(x + L)\}. \quad (30.8)$$

G , the group of actions $g \in G$ on a state space (reflections, translations, etc.) is a symmetry of the Kuramoto-Sivashinsky equation (30.2) if $g u_t = F(g u)$. A

continuous symmetry maps each state $u \in \mathbb{U}$ to a manifold of physically equivalent states. The Kuramoto-Sivashinsky equation is time translationally invariant, and space translationally invariant on a periodic domain under the 1-parameter group $O(2) = D_{1,x} \ltimes SO(2) : \{\sigma, \tau_{\ell/L}\}$. If $u(x, t)$ is a solution, then $\tau_{\ell/L} u(x, t) = u(x + \ell, t)$ is an equivalent solution for any shift $0 \leq \ell < L$, as is the reflection ('parity' or 'inversion')

$$\sigma u(x) = -u(-x). \quad (30.9)$$

Reflection generates the dihedral subgroup $D_1 = Z_2 = \{1, \sigma\}$ of $O(2)$. Relation $\sigma^2 = 1$ induces linear decomposition $u(x) = u^+(x) + u^-(x)$, $u^\pm(x) = P^\pm u(x) \in \mathbb{U}^\pm$, into irreducible subspaces $\mathbb{U} = \mathbb{U}^+ \oplus \mathbb{U}^-$, where

$$P^+ = (1 + \sigma)/2, \quad P^- = (1 - \sigma)/2, \quad (30.10)$$

are the antisymmetric/symmetric projection operators. Applying P^+ , P^- on the KS equation (30.2) we have

$$\begin{aligned} u_t^+ &= -(u^+ u_x^+ + u^- u_x^-) - u_{xx}^+ - u_{xxxx}^+ \\ u_t^- &= -(u^+ u_x^- + u^- u_x^+) - u_{xx}^- - u_{xxxx}^-. \end{aligned} \quad (30.11)$$

If $u^- = 0$, KS flow is confined to the antisymmetric $\mathbb{U}^+$ subspace,

$$u_t^+ = -u^+ u_x^+ - u_{xx}^+ - u_{xxxx}^+, \quad (30.12)$$

but otherwise the nonlinear terms in (30.11) mix the two subspaces.



example 30.1
p. 631



example 30.2
p. 631



example 30.4
p. 632

30.2 Constructing a state space

Spatial periodicity $u(x, t) = u(x + L, t)$ makes it convenient to work in the Fourier space,

$$u(x, t) = \sum_{k=-\infty}^{+\infty} a_k(t) e^{ikx/\tilde{L}}, \quad (30.13)$$

with the 1-dimensional PDE (30.2) replaced by an infinite set of ODEs for the complex Fourier coefficients $a_k(t)$:

$$\dot{a}_k = v_k(a) = (q_k^2 - q_k^4) a_k - i \frac{q_k}{2} \sum_{m=-\infty}^{+\infty} a_m a_{k-m}, \quad q_k = k/\tilde{L}. \quad (30.14)$$

As $\dot{a}_0 = 0$, a_0 is a conserved quantity, in our calculations fixed to $a_0 = 0$ by the vanishing mean $\langle u \rangle$ condition (30.3) for the front velocity. The velocity field $u(x, t)$ is real, so $a_k = a_{-k}^*$, and we can replace the sum by an $m > 0$ sum.

This is the infinite set of ordinary differential equations promised in this chapter's introduction.

As $a_0 = 0$ in (30.14), a_0 is a conserved quantity fixed to $a_0 = 0$ by the condition (30.3).

The translation operator action on the Fourier coefficients (2.19), represented here by a complex valued vector $a = \{a_k \in \mathbb{C} | k = 1, 2, \dots\}$, is given by

$$\tau_{\ell/L} a = \mathbf{g}(\ell) a, \quad (30.15)$$

where $\mathbf{g}(\ell) = \text{diag}(e^{iq_k \ell})$ is a complex valued diagonal matrix, which amounts to the k th mode complex plane rotation by an angle $k \ell / \tilde{L}$. The reflection acts on the Fourier coefficients by complex conjugation,

$$\sigma a = -a^*. \quad (30.16)$$

Due to the hyperviscous damping u_{xxxx} , long time solutions of Kuramoto-Sivashinsky equation are smooth, a_k drop off fast with k , and truncations of (30.14) to N terms, $16 \leq N \leq 128$, yield highly accurate solutions for system sizes considered here.

Which brings us to the single, universal, correct spatial mode-indexing convention for *all* Kuramoto-Sivashinsky papers you might plan to write up. It's our misfortune that a few amongst living have as much experience with Kuramoto-Sivashinsky as we do, so please spare us of the pain of converting your $L = 2\pi$, varying 'hyper-viscosity' spacetime plots into the physically meaningful, varying 'Rayleigh number' $\tilde{L}$ (30.4), correctly 'recoloring' field $u(x, t)$ plots. In our convention, small L -domains correctly overlay longer L -domains, and that's all there is to the only sensible convention.

In Fourier representations, always use units of $k/\tilde{L} = 2\pi k/L$ along the horizontal, eigenmode / wavenumber abscissa. This way the wave-number plots for any L share the same abscissa units; for example, the most unstable mode $a_{k_{max}}$, $k_{max}/\tilde{L} \approx \sqrt{2}$ is close the same abscissa value $\sqrt{2}$ for different system size L plots.

As Kuramoto-Sivashinsky has $O(2)$ symmetry, with its 2-dimensional irreducible representations, i.e., sines & cosines, rather than $\exp(ik/\tilde{L})$'s, always group Floquet exponents and Fourier modes into $(k, k+1)$ pairs, plot them as a single, two-valued function of $k/\tilde{L}$. What one chooses to pair for low k might be ambiguous, as the nonlinear interactions mix up the $O(2)$ 2-dimensional linearly irreducible representations.

You might also want to use logarithmic scale on the y-axis (ordinate), as modes are expected to fall off exponentially or faster for modes beyond the entangled, "physical", inertial-manifold modes. To trust your calculation, you always want to capture the beginning of the transient, "unphysical" modes in the spectrum; for example, modes beyond $k = 7$ in figure 30.8.

Should you know of any reason that we should not be joined in this holy notational convention, speak now or forever hold your peace.

Robustness of the Fourier representation of KS as a function of the number of modes kept in truncations is a subtle issue. Adding an extra mode to a truncation introduces a small perturbation. However, this can (and often will) throw the dynamics into a different asymptotic state. A chaotic attractor for $N = 15$ can collapse into an attractive period-3 cycle for $N = 16$, and so on. If we compute, for example, the Lyapunov exponent λ_N for a strange attractor of the system (30.14), there is no reason to expect λ_N to smoothly converge to a limit value λ , as $N \rightarrow \infty$, because of the lack of structural stability both as a function of truncation N , and the system size $\tilde{L}$. However, later in this chapter we explore both equilibria and short periodic orbits, which are robust under mode truncations and small system parameter $\tilde{L}$ changes. Spatial representations of PDEs (such as figure 30.1 (b) and the 3D snapshots of velocity and vorticity fields in Navier-Stokes) offer little insight into detailed dynamics of low- Re flows. Much more illuminating are the state space representations.

30.2.1 Equilibria and relative equilibria

Equilibria (or the steady solutions) are the fixed profile time-invariant solutions,

$$u(x, t) = u_q(x). \quad (30.17)$$

Due to the translational symmetry, the KS system also allows for relative equilibria (traveling waves, rotating waves), characterized by a fixed profile $u_q(x)$ moving with constant speed c , i.e.

$$u(x, t) = u_q(x - ct). \quad (30.18)$$

Here suffix q labels a particular invariant solution. Because of the reflection symmetry (30.9), the relative equilibria come in counter-traveling pairs $u_q(x - ct)$, $-u_q(-x + ct)$.

The set of equilibria, relative equilibria, and their stable / unstable manifolds are important for us, as they form the coarsest topological framework for organizing state space orbits.



example 30.6
p. 633



example 30.7
p. 634

In the Fourier representation the relative equilibria time dependence is

$$a_k(t)e^{-itcq_k} = a_k(0). \quad (30.19)$$

Differentiating with respect to time, we obtain the Fourier space version of the relative equilibrium condition (30.43),

$$v_k(a) - iq_k c a_k = 0, \quad (30.20)$$

which we solve for (time independent) a_k and c .

Periods of spatially periodic equilibria are multiples of L . Every time the system size crosses $\tilde{L} = n$, n -cell states are generated through pitchfork bifurcations

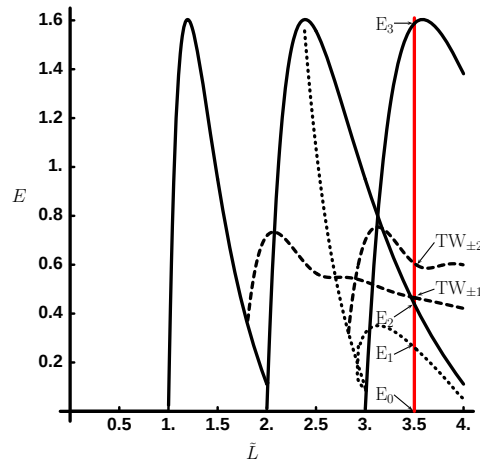


Figure 30.2: The energy (30.30) of the equilibria and relative equilibria that exist up to $L = 22$, $\tilde{L} = 3.5014 \dots$, plotted as a function of the system size $\tilde{L} = L/2\pi$ (additional equilibria, not present at $L = 22$ are given in ref. [21]). Solid curves denote 2- and 3-cell solutions E_2 and E_3 , dotted curves the GLMRT equilibrium E_1 , and dashed curves the relative equilibria $TW_{\pm 1}$ and $TW_{\pm 2}$. The parameter α of refs. [21, 27] is related to the system size by $\tilde{L} = \sqrt{\alpha/4}$.

off $u = 0$ equilibrium. Due to the translational invariance of Kuramoto-Sivashinsky equation, they form invariant circles in the full state space. In the $\mathbb{U}^+$ subspace considered here, they correspond to $2n$ points, each shifted by $L/2n$. For a sufficiently small L the number of equilibria is small and concentrated on the low wave-number end of the Fourier spectrum.

From (30.14) we see that the origin $u(x, t) = 0$ has Fourier modes as the linear stability eigenvectors. The $|k| < \tilde{L}$ long wavelength perturbations of the flat-front equilibrium are linearly unstable, while all $|k| > \tilde{L}$ short wavelength perturbations are strongly contractive. The high k eigenvalues, corresponding to rapid variations of the flame front, decay so fast that the corresponding eigendirections are physically irrelevant. The most unstable mode, nearest to $|k| = \tilde{L}/\sqrt{2}$, sets the scale of the mean wavelength $\sqrt{2}$ of the KS ‘turbulent’ dynamics, see figure 30.1.

30.2.2 Relative periodic orbits, symmetries and periodic orbits

The KS equation (30.2) is time translationally invariant, and space translationally invariant under the 1- d Lie group of $O(2)$ rotations: if $u(x, t)$ is a solution, then $u(x + \ell, t)$ and $-u(-x, t)$ are equivalent solutions for any $-L/2 < \ell \leq L/2$. As a result of invariance under $\tau_{\ell/L}$, KS equation can have relative periodic orbit solutions with a profile $u_p(x)$, period T_p , and a nonzero shift ℓ_p

$$\tau_{\ell_p/L} u(x, T_p) = u(x + \ell_p, T_p) = u(x, 0) = u_p(x). \quad (30.21)$$

Relative periodic orbits (30.21) are periodic in $v_p = \ell_p/T_p$ co-rotating frame (see figure 12.8), but in the stationary frame their trajectories are quasiperiodic. Due to the reflection symmetry (30.9) of KS equation, every relative periodic orbit $u_p(x)$ with shift ℓ_p has a symmetric partner $-u_p(-x)$ with shift $-\ell_p$.

Due to invariance under reflections, KS equation can also have relative periodic orbits *with reflection*, which are characterized by a profile $u_p(x)$ and period T_p

$$\sigma u(x + \ell, T_p) = -u(-x - \ell, T_p) = u(x + \ell, 0) = u_p(x), \quad (30.22)$$

giving the family of equivalent solutions parameterized by ℓ (as the choice of the reflection point is arbitrary, the shift can take any value in $-L/2 < \ell \leq L/2$).

Armbruster *et al.* [1, 2] and Brown and Kevrekidis [6] (see also ref. [28]) link the birth of relative periodic orbits to an infinite period global bifurcation involving a heteroclinic loop connecting equilibria or a bifurcation of relative equilibria, and also report creation of relative periodic orbit branches through bifurcation of periodic orbits.

As ℓ is continuous in the interval $[-L/2, L/2]$, the likelihood of a relative periodic orbit with $\ell_p = 0$ shift is zero, unless an exact periodicity is enforced by a discrete symmetry, such as the dihedral symmetries discussed above. If the shift ℓ_p of a relative periodic orbit with period T_p is such that ℓ_p/L is a rational number, then the orbit is periodic with period nT_p . The likelihood to find such periodic orbits is also zero.

However, due to the KS equation invariance under the dihedral D_n and cyclic C_n subgroups, the following types of periodic orbits are possible:

(a) The periodic orbit lies within a subspace pointwise invariant under the action of D_n or C_n . For instance, for D_1 this is the $\mathbb{U}^+$ antisymmetric subspace, $-u_p(-x) = u_p(x)$, and $u(x, T_p) = u(x, 0) = u_p(x)$. The periodic orbits found in refs. [9, 32] are all in $\mathbb{U}^+$, as the dynamics is restricted to antisymmetric subspace. For $L = 22$ the dynamics in $\mathbb{U}^+$ is dominated by attracting (within the subspace) heteroclinic connections and thus we have no periodic orbits of this type, or in any other of the D_n -invariant subspaces.

(b) a pre-periodic orbit satisfies

$$u(x, t + T_p) = \gamma u(x, t), \quad (30.23)$$

for some group element $\gamma \in O(2)$ such that $\gamma^m = e$ for some integer m so that the orbit repeats after time mT_p (see ref. [20] for a general discussion of conditions on the symmetry of periodic orbits). If an orbit is of reflection type (30.22), $\sigma \tau_{\ell/L} u(x, T_p) = -u(-x - \ell, T_p) = u(x, 0)$, then it is pre-periodic to a periodic orbit with period $2T_p$. Indeed, since $(\sigma \tau_{\ell/L})^2 = \sigma^2 = 1$, and the KS solutions are time translation invariant, it follows from (30.22) that

$$u(x, 2T_p) = \sigma \tau_{\ell/L} u(x, T_p) = (\sigma \tau_{\ell/L})^2 u(x, 0) = u(x, 0).$$

Thus any shift acquired during time 0 to T_p is compensated by the opposite shift during evolution from T_p to $2T_p$. All periodic orbits we have found for $L = 22$ are of type (30.23) with $\gamma = R$. Pre-periodic orbits with $\gamma \in C_n$ have been found by Brown and Kevrekidis [6] for KS system sizes larger than ours, but we have not found any for $L = 22$. Pre-periodic orbits are a hallmark of any dynamical system with a discrete symmetry, where they have a natural interpretation as periodic orbits in the fundamental domain.

30.3 Energy budget

Mathematical physics is three things: Gaussian integrals, integration by parts and ... (nobody remembers exactly what the third thing was, including Joel).

—Joel Lebowitz, in a seminar

In physical settings where the observation times are much longer than the dynamical ‘turnover’ and Lyapunov times (statistical mechanics, quantum physics, turbulence) periodic orbit theory provides highly accurate predictions of measurable long-time averages such as the dissipation and the turbulent drag. Physical predictions have to be independent of a particular choice of ODE representation of the PDE under consideration and invariant under all symmetries of the dynamics. In this section we discuss a set of such physical observables for the 1- d KS invariant under reflections and translations. Here we shall show that they offer a visualization of solutions of dynamics in which the symmetries are explicitly quotiented out.

The *space average* of a function $a = a(x, t) = a(u(x, t))$ periodic on the interval L is given by

$$\langle a \rangle = \frac{1}{L} \oint dx a(x, t), \quad (30.24)$$

We note that total derivatives vanish by the spatial periodicity on the L domain, and that by integration by parts

$$\langle f_x \rangle = 0, \quad \langle f_x g \rangle = -\langle f g_x \rangle \quad (30.25)$$

for any L -periodic functions f, g . In general $\langle a \rangle$ is time dependent. Its *mean value* is given by the *Birkhoff average*

$$\bar{a} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t d\tau \langle a \rangle = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \frac{1}{L} \oint d\tau dx a(x, \tau). \quad (30.26)$$

The mean value of $a_q = a(u_q)$ evaluated on equilibrium or relative equilibrium $u(x, t) = u_q(x - ct)$, is

$$\bar{a}_q = \langle a \rangle_q = a_q. \quad (30.27)$$

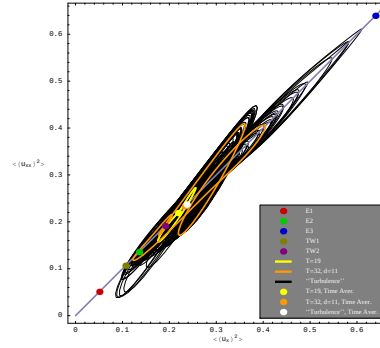
Evaluation of the infinite time average (30.26) on a function of a periodic orbit or relative periodic orbit $u_p(x, t) = u_p(x + \ell_p, t + T_p)$ requires only a single T_p traversal,

$$\bar{a}_p = \frac{1}{T_p} \int_0^{T_p} d\tau \langle a \rangle. \quad (30.28)$$

Equation (30.2) can be written as

$$u_t = -V_x, \quad V(x, t) = \frac{1}{2}u^2 + u_x + u_{xxx}, \quad (30.29)$$

Figure 30.3: Power input $\langle u_x^2 \rangle$ vs. dissipation $\langle u_{xx}^2 \rangle$ for $L = 22$ equilibria and relative equilibria, for several periodic orbits and relative periodic orbits, and for a typical ‘turbulent’ state. Note that $\overline{(u_{p,x})^2}$ of the $(T_p, \ell_p) = (32.8, 10.96)$ relative periodic orbit, which appears well embedded within the turbulent state, is close to the turbulent expectation $\overline{(u_x)^2}$ (from ref. [11]).



and E can be interpreted as the mean energy density (30.30). So, even though KS is a phenomenological small-amplitude equation, the time-dependent L^2 norm of u (simplified using integration by parts, as in (30.25)),

$$E = \frac{1}{L} \oint dx V(x, t) = \frac{1}{L} \oint dx \frac{u^2}{2}, \quad (30.30)$$

has a physical interpretation as the average ‘energy’ density of the flame front. This analogy to the mean kinetic energy density for the Navier-Stokes motivates what follows.

The energy (30.30) is intrinsic to the flow, independent of the particular ODE basis set chosen to represent the PDE. As the Fourier amplitudes are eigenvectors of the translation operator, in the Fourier space the energy is a diagonalized quadratic norm,

$$E = \sum_{k=1}^{\infty} E_k, \quad E_k = \frac{1}{2} |a_k|^2, \quad (30.31)$$

and explicitly invariant term by term under translations and reflections.

Take time derivative of the energy density (30.30), substitute (30.2) and integrate by parts, as in (30.25):

$$\begin{aligned} \dot{E} &= \langle u_t u \rangle = -\langle (u^2/2 + u_x + u_{xxx})_x u \rangle \\ &= \langle u_x u^2/2 + u_x^2 + u_x u_{xxx} \rangle. \end{aligned} \quad (30.32)$$

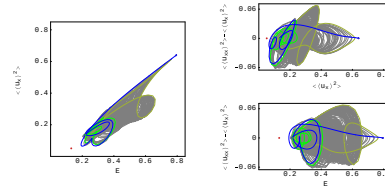
The first term in (30.32) vanishes by integration by parts, $3\langle u_x u^2 \rangle = \langle (u^3)_x \rangle = 0$, and integrating the third term by parts yet again one gets that the energy variation in the Kuramoto-Sivashinsky equation (30.2)

$$\dot{E} = P - D, \quad P = \langle u_x^2 \rangle, \quad D = \langle u_{xx}^2 \rangle \quad (30.33)$$

balances the power P pumped in by anti-diffusion u_{xx} against the energy dissipation rate D by hyper-viscosity u_{xxx} .

In figure 30.3 we plot the power input $\langle u_x^2 \rangle$ vs. dissipation $\langle u_{xx}^2 \rangle$ for all $L = 22$ equilibria and relative equilibria determined so far, several periodic orbits

Figure 30.4: EQ_1 (red), EQ_2 (green), EQ_3 (blue), connections from EQ_1 to $A(L/4)EQ_1$ (green), from $A(L/4)EQ_1$ to EQ_1 (yellow-green) and from EQ_3 to $A(L/4)EQ_1$ (blue), along with a generic long-time “turbulent” evolution (grey) for $L = 22$. Three different projections of the $(E, \langle u_x^2 \rangle, \langle u_{xx}^2 \rangle - \langle u_x^2 \rangle)$ representation are shown (from ref. [11]).



and relative periodic orbits, and for a typical “turbulent” evolution. The Birkhoff averaged energy density $\bar{E}$ computed on a typical orbit goes to a constant, so the mean values (30.26) of drive and dissipation exactly balance each other:

$$\bar{\dot{E}} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t d\tau \dot{E} = \bar{P} - \bar{D} = 0. \quad (30.34)$$

In particular, the equilibria and relative equilibria fall onto the diagonal in figure 30.3 (a), and so do Birkhoff averages computed on periodic orbits and relative periodic orbits:

$$E_p = \frac{1}{T_p} \int_0^{T_p} d\tau E(\tau), \quad P_p = \frac{1}{T_p} \int_0^{T_p} d\tau P(\tau) = \bar{D}_p. \quad (30.35)$$

In the Fourier basis (30.31) the conservation of energy on average takes form

$$0 = \sum_{k=-\infty}^{\infty} (q_k^2 - q_k^4) \bar{E}_k, \quad E_k(t) = \frac{1}{2} |a_k(t)|^2. \quad (30.36)$$

The large k convergence of this series is insensitive to the system size L ; $\bar{E}_k$ have to decrease much faster than q_k^{-4} . Deviation of E_k from this bound for small k determines the active modes. This may be useful to bound the number of equilibria, with the upper bound given by zeros of a small number of long wavelength modes.

30.4 Infinite-dimensional flows: Numerics

The computer is not a mere mathematical excrescence, useful for technological ends. Rather, I believe that it is a meta-development that might very well change what mathematics is considered to be.

— P. J. Davis [13]

The trivial solution $u(x, t) = 0$ is an equilibrium point of (30.2), but that is basically all we know as far as useful analytical solutions are concerned. To develop some intuition about the dynamics we turn to numerical simulations.

How are solutions such as figure 30.1 computed? The salient feature of such partial differential equations is a theorem saying that for state space contracting

Figure 30.5: Spatiotemporally periodic solution $u_0(x, t)$, with period $T_0 = 30.0118$. The antisymmetric subspace, $u(x, t) = -u(-x, t)$, so we plot $x \in [0, L/2]$. System size $\tilde{L} = 2.89109$, $N = 16$ Fourier modes truncation (from ref. [9]).

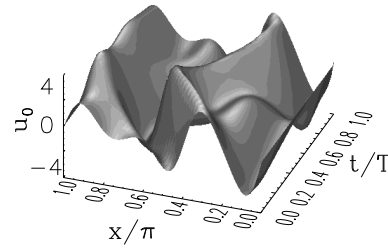
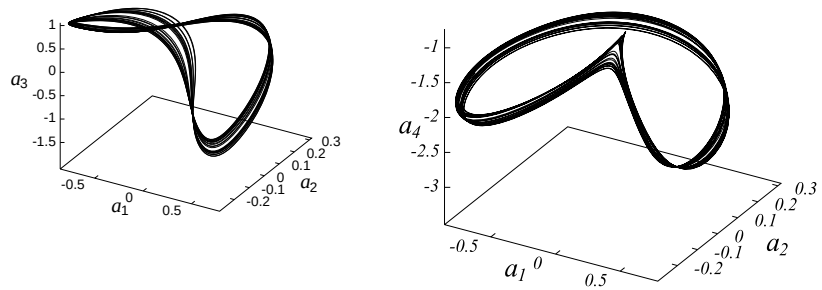


Figure 30.6: Projections of a typical 16-dimensional trajectory onto different 3-dimensional subspaces, coordinates (a) $\{a_1, a_2, a_3\}$, (b) $\{a_1, a_2, a_4\}$. System size $\tilde{L} = 2.89109$, $N = 16$ Fourier modes truncation (from ref. [9]).



flows, the asymptotic dynamics is describable by a *finite* set of ‘inertial manifold’ ordinary differential equations. How you solve the equation (30.2) numerically is up to you. Here are some options:

Discrete mesh: You can divide the x interval into a sufficiently fine discrete grid of N points, replace space derivatives in (30.2) by approximate discrete derivatives, and integrate a finite set of first order differential equations for the discretized spatial components $u_j(t) = u(jL/N, t)$, by any integration routine you trust.

Fourier modes: You can integrate numerically the Fourier modes (30.14), truncating the ladder of equations to a finite number of modes N , i.e., set $a_k = 0$ for $k > N$. In the applied mathematics literature more sophisticated variants of such truncations are called *Galerkin truncations*, or *Galerkin projections*. You need to worry about ‘stiffness’ of the equations and the stability of your integrator. For the parameter values explored in this chapter, truncations N in range 16 to 64 yield sufficient accuracy.

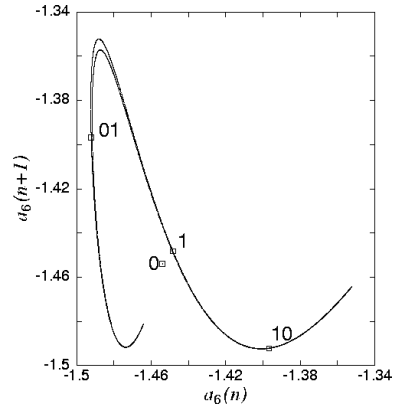
exercise 2.6

Pseudo-spectral methods: You can mix the two methods, exploiting the speed of Fast Fourier Transforms.



example 30.3
p. 632

Figure 30.7: The attractor of the Kuramoto-Sivashinsky system (30.14), plotted as the a_6 component of the $a_1 = 0$ Poincaré section return map. Here 10,000 Poincaré section returns of a typical trajectory are plotted. Also indicated are the periodic points 0, 1, 01 and 10. System size $\tilde{L} = 2.89109$, $N = 16$ Fourier modes truncation (from ref. [9]).



30.5 Visualization

The ultimate goal, however, must be a rational theory of statistical hydrodynamics where [...] properties of turbulent flow can be mathematically deduced from the fundamental equations of hydromechanics.

—E. Hopf

The problem with high-dimensional representations, such as truncations of the infinite tower of equations (30.14), is that the dynamics is difficult to visualize. The best we can do without much programming is to examine the trajectory's projections onto any three axes a_i, a_j, a_k , as in figure 30.6.

example 30.4

The question is: how is one to look at such a flow? It is not clear that restricting the dynamics to a Poincaré section necessarily helps - after all, a section reduces a $(d + 1)$ -dimensional flow to a d -dimensional map, and how much is gained by replacing a continuous flow in 16 dimensions by a set of points in 15 dimensions? The next example illustrates the utility of visualization of dynamics by means of Poincaré sections.

The example 30.4 illustrates why a Poincaré section gives a more informative snapshot of the flow than the full flow portrait. While no fine structure is discernible in the full state space flow portraits of the Kuramoto-Sivashinsky dynamics, figure 30.6, the return map figure 30.7 reveals the fractal structure in the asymptotic attractor.

In order to find a better representation of the dynamics, we now turn to its topological invariants.

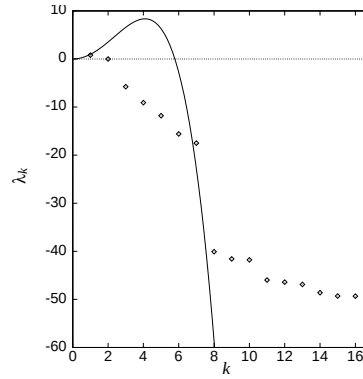
30.6 Why does a flame front flutter?

I understood every word.

—Fritz Haake

section 21.2

Figure 30.8: Floquet exponents λ_k (5.4) versus k for the least unstable spatio-temporally periodic orbit $\bar{\Gamma}$ of the Kuramoto-Sivashinsky system, compared with the stability exponents (5.2) of the $u(x, t) = 0$ flat-front equilibrium, $\lambda_k = k^2 - k^4$. The eigenvalues λ_k for $k \geq 8$ fall below the numerical accuracy of integration and are not meaningful. The cycle $\bar{\Gamma}$ was computed using methods of chapter 16. System size $\tilde{L} = 2.89109$, $N = 16$ Fourier modes truncation (from ref. [9]).



We start by considering the case where a_q is an equilibrium point (2.11). Expanding around the equilibrium point a_q , and using the fact that the matrix $A = A(a_q)$ in (4.2) is constant, we can apply the simple formula (5.1) also to the Jacobian matrix of an equilibrium point of a PDE,

$$J'(a_q) = e^{At} \quad A = A(a_q).$$

For $\tilde{L} < 1$, $u(x, t) = 0$ is the globally attractive stable equilibrium. As the system size $\tilde{L}$ is increased, the “flame front” becomes increasingly unstable and turbulent, the dynamics goes through a rich sequence of bifurcations sketched in figure 30.2 that which we shall not dwell on here.

According to (30.6) the $|k| < \tilde{L}$ long wavelength perturbations of the flat-front equilibrium are linearly unstable, while all $|k| > \tilde{L}$ short wavelength perturbations are strongly contractive. The high k eigenvalues, corresponding to rapid variations of the flame front, decay so fast that the corresponding eigen-directions are physically irrelevant. To illustrate the rapid contraction in the non-leading eigen-directions we plot in figure 30.8 the eigenvalues of the equilibrium in the unstable regime, for relatively small system size, and compare them with the Floquet multipliers of the least unstable cycle for the same system size. The equilibrium solution is very unstable, in 5 eigen-directions, the least unstable cycle only in one. Note that for $k > 7$ the rate of contraction is so strong that higher eigen-directions are numerically meaningless for either solution; even though the flow is infinite-dimensional, the attracting set must be rather thin.

While in general for $\tilde{L}$ sufficiently large one expects many coexisting attractors in the state space, in numerical studies most random initial conditions seem to settle on the same chaotic attractor.

From (30.14) we see that the $u(x, t) = 0$ flat-front equilibrium has Fourier modes as the linear stability eigenvectors. For $|k| < \tilde{L}$, the corresponding Fourier modes are unstable. The most unstable mode has $k = \tilde{L}/\sqrt{2}$ and defines the scale of basic building blocks of the spatiotemporal dynamics of the Kuramoto-Sivashinsky equation in large system size limit.

Consider now the case of initial a_k sufficiently small that the bilinear $a_m a_{k-m}$ terms in (30.14) can be neglected. Then we have a set of decoupled linear equations for a_k whose solutions are exponentials, at most a finite number for which

$|k| \leq \tilde{L}$ is growing with time, and infinitely many with $|k| > \tilde{L}$ decaying in time. The growth of the unstable long wavelengths (low $|k|$) excites the short wavelengths through the $a_m a_{k-m}$ nonlinear term. The excitations thus transferred are dissipated by the strongly damped short wavelengths, and a “chaotic equilibrium” can emerge. The very short wavelengths $|k| \gg 1/\sqrt{\nu}$ remain small for all times, but the intermediate wavelengths of order $|k| \sim 1/\sqrt{\nu}$ play an important role in maintaining the dynamical equilibrium. As the damping parameter decreases, the solutions increasingly take on shock front character poorly represented by the Fourier basis, and many higher harmonics may need to be kept in truncations of (30.14).

Hence, while one may truncate the high modes in the expansion (30.14), care has to be exercised to ensure that no modes essential to the dynamics are chopped away.

In other words, even though our starting point (30.2) is an infinite-dimensional dynamical system, the asymptotic dynamics unfolds on a finite-dimensional attracting manifold, and so we are back on the familiar territory of sect. 2.2: the theory of a finite number of ODEs applies to this infinite-dimensional PDE as well.

We can now start to understand the remark on page 45 that for infinite dimensional systems time reversibility is not an option: evolution forward in time strongly damps the higher Fourier modes. There is no turning back: if we reverse the time, the infinity of high modes that contract strongly forward in time now explodes, instantly rendering evolution backward in time meaningless. As so much you are told about dynamics, this claim is also wrong, in a subtle way: if the initial $u(x, 0)$ is in the non-wandering set (2.5), the trajectory is well defined both forward and backward in time. For practical purposes, this subtlety is not of much use, as any time-reversed numerical trajectory in a finite-mode truncation will explode very quickly, unless special precautions are taken.

When is an equilibrium important? There are two kinds of roles equilibria play:

“Hole” in the natural measure. The more unstable eigen-directions it has (for example, the $u = 0$ solution), the more unlikely it is that an orbit will recur in its neighborhood.

Unstable manifold of a “least unstable” equilibrium. Asymptotic dynamics spends a large fraction of time in neighborhoods of a few equilibria with only a few unstable eigen-directions.

30.7 Intrinsic parametrization

Both in the Rössler flow of example 3.2, and in the Kuramoto-Sivashinsky system of example 30.4 we have learned that the attractor is very thin, but otherwise the

Figure 30.9: The return map of the Kuramoto-Sivashinsky system (30.14) figure 30.7, from the unstable manifold of the $\bar{1}$ fixed point to the (neighborhood of) the unstable manifold. Also indicated are the periodic points $\bar{0}$ and $\bar{0}\bar{1}$.

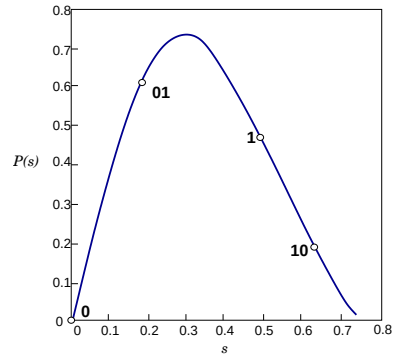
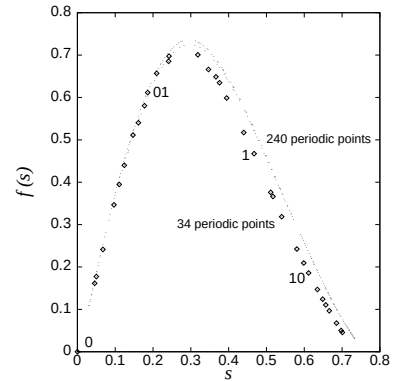


Figure 30.10: The return map $s_{n+1} = f(s_n)$ constructed from the images of periodic points. The diamonds were obtained by using 34 periodic points, and the tiny dots were obtained by using 240 periodic points. We have indicated the periodic points $\bar{0}$, $\bar{1}$ and $\bar{0}\bar{1}$. Note that the transverse fractal structure of the map shows when the number of points is increased. System size $\tilde{L} = 2.89109$, $N = 16$ Fourier modes truncation (from ref. [9]).



return maps that we found were disquieting – neither figure 3.4 nor figure 30.7 appeared to be one-to-one maps. This apparent loss of invertibility is an artifact of projection of higher-dimensional return maps onto lower-dimensional subspaces. As the choice of lower-dimensional subspace is arbitrary, the resulting snapshots of return maps look rather arbitrary, too. Other projections might look even less suggestive.

Such observations beg a question: Does there exist a ‘natural’, intrinsically optimal coordinate system in which we should plot of a return map?

As we shall now argue (see also sect. 16.1), the answer is yes: The intrinsic coordinates are given by the stable/unstable manifolds, and a return map should be plotted as a map from the unstable manifold back onto the immediate neighborhood of the unstable manifold.

Examination of numerical plots such as figure 30.6 suggests that a more thoughtful approach would be to find a coordinate transformation $y = h(x)$ to a ‘center manifold’, such that in the new, curvilinear coordinates large-scale dynamics takes place in (y_1, y_2) coordinates, with exponentially small dynamics in $y_3, y_4 \dots$. But - thinking is extra price - we do not know how to actually accomplish this, and we do not believe it can be accomplished globally.

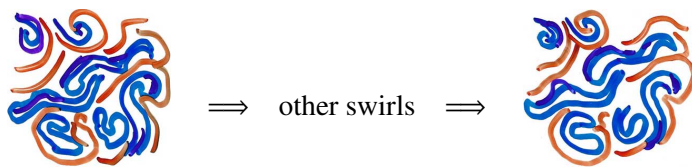
Both in the example of the Rössler flow and of the Kuramoto-Sivashinsky system we sketched the attractors by running a long chaotic trajectory, and noted that the attractors are very thin, but otherwise the return maps that we plotted were disquieting – neither figure 3.4 nor figure 30.7 appeared to be 1-to-1 maps. In this section we show how to use such information to approximately locate cycles.

Résumé

Turbulence is the graveyard of theories
— Hans W. Liepmann

We have learned that an instanton is an analytic solution of Yang-Mills equations of motion, but shouldn't a strongly nonlinear field theory dynamics be dominated by turbulent solutions? How are we to think about systems where every spatiotemporal solution is unstable?

Here we think of turbulence in terms of *recurrent* spatiotemporal patterns. Pictorially, dynamics drives a given spatially extended system through a repertoire of unstable patterns; as we watch a turbulent system evolve, every so often we catch a glimpse of a familiar pattern:



For a finite spatial resolution and a finite time, a pattern belonging to a finite alphabet of admissible patterns is observed; the long term dynamics can be thought of as a walk through the space of such patterns. Recasting this image into mathematics is what ChaosBook is about.

The problem one faces with high-dimensional flows is that their topology is hard to visualize, and that even with a decent starting guess for a point on a periodic orbit, methods like the Newton-Raphson method are likely to fail. Methods that start with initial guesses for a number of points along the cycle, such as the multipoint shooting method of sect. 16.2, are more robust. The relaxation (or variational) methods take this strategy to its logical extreme, and start by a guess of not a few points along a periodic orbit, but a guess of the entire orbit. As these methods are intimately related to variational principles and path integrals, we postpone their introduction to chapter 34.

chapter 34

At present the theory is in practice applicable only to systems with a low *intrinsic dimension* – the minimum number of coordinates necessary to capture its essential dynamics. If the system is very turbulent (a description of its long time dynamics requires a space of very high intrinsic dimension) we are out of luck.

Commentary

In saying this, I imply no criticism of the Court, which in those cases was faced with the task of trying to define what may be indefinable. [...] I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it, [...]

—Justice Potter Stewart, *Jacobellis v. Ohio* (1964)

Remark 30.1. A brief history of dynamicist's vision of turbulence. Dynamical approaches to study of turbulence are - surprisingly - still a cutting-edge research area. We have not even agreed yet on when one is allowed to use word 'turbulence'. For many practitioners, a wide range of scales (i.e., high Reynolds numbers) is the essential ingredient of turbulence, with intermittency, fractality, ..., possibly exhibited by a particular turbulent flow, but not essential. When and how ChaosBook uses words 'chaos', 'spatiotemporal chaos' and 'turbulence' is explained in sect. 1.3.2 *What is 'turbulence'?*. You might find also Appendix A1.5 amusing.

appendix A1.5

The precise definition of the Reynolds number depends on the system, the boundary conditions and the forcing terms. Once that is fixed, one can compute bifurcations and determine the onset of chaos to a desired precision. Some would claim we may only call it 'turbulence' when there is an inertial range, but there is mounting evidence, by masters of fully developed turbulent flows and large eddy simulations [8, 40], that transitional turbulence contains the key ingredients of fully developed turbulence, and that changes in the flow as the Reynolds number increases are continuous [3, 4, 8, 40]. So, when does the multi-scale character start? Even at the onset of chaos one observes a broad spatial Fourier spectrum, indicating many active scales. Then, as the Reynolds number is increased, progressively smaller scales become active, whereas the large remain. If, in your opinion, what we do is not 'turbulence', please pinpoint at what Reynolds number turbulence, as defined by you, begins and what happens exactly then.

The work described in this chapter was initiated by Vakhtang Putkaradze's 1996 ChaosBook term project (see ChaosBook.org/extras), and continued by Budanur, Christiansen, Cvitanović, Davidchack, Ding, Gudorf, Lan, and Siminos [7, 9–12, 14–16, 19, 22, 23, 30–32]. You too can initiate a whole full-fledged research program by carrying out a good ChaosBook course project.

For a clear and simple overview of how to compute periodic orbits in higher dimensions (for example, for Navier-Stokes), see Willis [39] lectures on *Equilibria, periodic orbits and computing them*, [arXiv:1908.06730](https://arxiv.org/abs/1908.06730), a very nice, student friendly introduction.

P. Cvitanović and M. Avila

Remark 30.2. Visualizations of steady turbulence. Visualizations of a high-dimensional state space trajectory are of necessity projections onto two or three dimensions. An appealing choice is to monitor the flow in terms of physical, symmetry-invariant observables, measured in units of their laminar values. In fluid dynamics *Reynolds-Orr energy equation* [36, 37] or *power-dissipation balance equation* for instantaneous kinetic energy $E(t)$, $\dot{E} = I - D$, balances the instantaneous power $I(t)$ pumped in the system against the energy dissipation rate $D(t)$ due to viscous effects. Greene and Kim [21] derive the corresponding equation for Kuramoto-Sivashinsky system (further physical observables beyond $(E(t), D(t), I(t))$ are difficult to construct). While such visualizations are common

in literature [26], projections onto a plane spanned by energy production and dissipation are misleading. As on average $\dot{E} \approx 0$, in such projections the ∞ -dimensional state space ergodic trajectory tends to stay close to the $I = D$ line, and much important information is lost: if two fluid states are clearly separated in such plot, they are also separated in the high-dimensional state space, but converse is not true; physically distinct states might have comparable dissipation rate, and such plots often obscure some of the most relevant features of the flow; more detailed examinations of the flow necessitate state space projections tailored to specific flow structures [11, 19].

The theorem on finite dimensionality of inertial manifolds of state space contracting PDE flows is proven in ref. [17]. Physical interpretation of the average “energy” density of the flame front of sect. 30.3, in analogy with the mean kinetic energy density for the Navier-Stokes, comes from Greene and Kim [21], a recommended reading. For equilibria the L -independent bound on E is given by Michelson [34]. The best current bound [5, 18] on the long-time limit of E as a function of the system size L scales as $E \propto L^{3/2}$, but deep in their hearts physicists know that E is extensive [15, 41], i.e., it scales linearly with the system size $E \propto L$.

Remark 30.3. Kuramoto-Sivashinsky. The Kuramoto-Sivashinsky equation was introduced in refs. [25, 29, 35, 38]. While it is usually written down in the “derivative” form (30.4), some authors prefer the “integral” form

$$h_t + h_{xx} + h_{xxxx} + \frac{1}{2}h_x^2 = 0, \quad \text{where } u = h_x, \quad (30.37)$$

see, for example, Michelson [34]. Holmes, Lumley and Berkooz [24] offer a delightful discussion of why this system deserves study as a staging ground for studying turbulence in full-fledged Navier-Stokes equation. How good a description of a flame front this equation is not a concern here; suffice it to say that such model amplitude equations for interfacial instabilities arise in a variety of contexts - see e.g. ref. [27] and EncyclopediaOfMath.org - and this one is perhaps the simplest physically interesting spatially extended nonlinear system.

Our criterion for reliable truncations of the infinite ladder of ordinary differential equations (30.14) is as follows. Adding an extra dimension to a truncation of the system (30.14) introduces a small perturbation, and this can (and often will) throw the system into a totally different asymptotic state. A chaotic attractor for $N = 15$ can become a period three window for $N = 16$, and so on. If we compute, for example, the Lyapunov exponent $\lambda(v, N)$ for the strange attractor of the system (30.14), there is no reason to expect $\lambda(v, N)$ to smoothly converge to the limit value $\lambda(v, \infty)$ as $N \rightarrow \infty$. The situation is different in the periodic windows, where the system is structurally stable, and it makes sense to compute Lyapunov exponents, escape rates, etc. for the *repeller*, i.e., the closure of the set of all *unstable* periodic orbits. Here the power of cycle expansions comes in: to compute quantities on the repeller by direct averaging methods is generally more difficult, because the asymptotic motion collapses to the stable cycle.

References

- [1] D. Armbruster, J. Guckenheimer, and P. Holmes, “[Heteroclinic cycles and modulated travelling waves in systems with O\(2\) symmetry](#)”, *Physica D* **29**, 257–282 (1988).

- [2] D. Armbruster, J. Guckenheimer, and P. Holmes, “Kuramoto-Sivashinsky dynamics on the center-unstable manifold”, *SIAM J. Appl. Math.* **49**, 676–691 (1989).
- [3] M. Avila and B. Hof, “Nature of laminar-turbulence intermittency in shear flows”, *Phys. Rev. E* **87**, 063012 (2013).
- [4] D. Barkley, B. Song, V. Mukund, G. Lemoult, M. Avila, and B. Hof, “The rise of fully turbulent flow”, *Nature* **526**, 550–553 (2015).
- [5] J. C. Bronski and T. N. Gambill, “Uncertainty estimates and L_2 bounds for the Kuramoto-Sivashinsky equation”, *Nonlinearity* **19**, 2023–2039 (2006).
- [6] H. S. Brown and I. G. Kevrekidis, “Modulated traveling waves for the Kuramoto-Sivashinsky equation”, in *Pattern Formation: Symmetry Methods and Applications*, edited by J. Chadam, M. Golubitsky, W. Langford, and B. Wetton (AMS, Providence, RI, 1996), pp. 45–66.
- [7] N. B. Budanur, *Exact Coherent Structures in Spatiotemporal Chaos: From Qualitative Description to Quantitative Predictions*, PhD thesis (School of Physics, Georgia Inst. of Technology, Atlanta, 2015).
- [8] R. T. Cerbus, C.-c. Liu, G. Gioia, and P. Chakraborty, “Small-scale universality in the spectral structure of transitional pipe flows”, *Sci. Adv.* **6**, eaaw6256 (2020).
- [9] F. Christiansen, P. Cvitanović, and V. Putkaradze, “Hopf’s last hope: Spatiotemporal chaos in terms of unstable recurrent patterns”, *Nonlinearity* **10**, 55–70 (1997).
- [10] P. Cvitanović, “Chaotic Field Theory: A sketch”, *Physica A* **288**, 61–80 (2000).
- [11] P. Cvitanović, R. L. Davidchack, and E. Siminos, “On the state space geometry of the Kuramoto-Sivashinsky flow in a periodic domain”, *SIAM J. Appl. Dyn. Syst.* **9**, 1–33 (2010).
- [12] P. Cvitanović and Y. Lan, Turbulent fields and their recurrences, in *Correlations and Fluctuations in QCD : Proceedings of 10. International Workshop on Multiparticle Production*, edited by N. Antoniou (2003), pp. 313–325.
- [13] P. J. Davis, “Spanning multiple worlds”, *SIAM News* **41** (2008).
- [14] X. Ding, *Geometry of Inertial Manifolds in Nonlinear Dissipative Dynamical Systems*, PhD thesis (School of Physics, Georgia Inst. of Technology, Atlanta, 2017).
- [15] X. Ding, H. Chaté, P. Cvitanović, E. Siminos, and K. A. Takeuchi, “Estimating the dimension of the inertial manifold from unstable periodic orbits”, *Phys. Rev. Lett.* **117**, 024101 (2016).
- [16] X. Ding and P. Cvitanović, “Periodic eigendecomposition and its application in Kuramoto-Sivashinsky system”, *SIAM J. Appl. Dyn. Syst.* **15**, 1434–1454 (2016).
- [17] C. Foias, M. S. Jolly, I. G. Kevrekidis, G. R. Sell, and E. S. Titi, “On the computation of inertial manifolds”, *Phys. Lett. A* **131**, 433–436 (1988).

- [18] L. Giacomelli and F. Otto, “New bounds for the Kuramoto-Sivashinsky equation”, *Commun. Pure Appl. Math.* **58**, 297–318 (2005).
- [19] J. F. Gibson, J. Halcrow, and P. Cvitanović, “Visualizing the geometry of state-space in plane Couette flow”, *J. Fluid Mech.* **611**, 107–130 (2008).
- [20] M. Golubitsky and I. Stewart, *The Symmetry Perspective* (Birkhäuser, Boston, 2002).
- [21] J. M. Greene and J.-S. Kim, “The steady states of the Kuramoto-Sivashinsky equation”, *Physica D* **33**, 99–120 (1988).
- [22] M. Gudorf and P. Cvitanović, Spatiotemporal tiling of the Kuramoto-Sivashinsky flow, In preparation, 2019.
- [23] M. N. Gudorf, *Spatiotemporal formulation of the Kuramoto-Sivashinsky equation*, PhD thesis (School of Physics, Georgia Inst. of Technology, Atlanta, 2019).
- [24] P. Holmes, J. L. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge Univ. Press, Cambridge UK, 1996).
- [25] G. M. Homsy, “Model equations for wavy viscous film flow”, in *Nonlinear Wave Motion*, Vol. 15, edited by A. C. Newell, Lectures in Applied Mathematics (Amer. Math. Soc., Providence, R.I, 1974), pp. 191–194.
- [26] G. Kawahara and S. Kida, “Periodic motion embedded in plane Couette turbulence: Regeneration cycle and burst”, *J. Fluid Mech.* **449**, 291–300 (2001).
- [27] I. G. Kevrekidis, B. Nicolaenko, and J. C. Scovel, “Back in the saddle again: a computer assisted study of the Kuramoto-Sivashinsky equation”, *SIAM J. Appl. Math.* **50**, 760–790 (1990).
- [28] M. Krupa, “Bifurcations of relative equilibria”, *SIAM J. Math. Anal.* **21**, 1453–1486 (1990).
- [29] Y. Kuramoto and T. Tsuzuki, “Persistent propagation of concentration waves in dissipative media far from thermal equilibrium”, *Progr. Theor. Phys.* **55**, 365–369 (1976).
- [30] Y. Lan, *Dynamical Systems Approach to 1 – d Spatiotemporal Chaos – A Cyclist’s View*, PhD thesis (School of Physics, Georgia Inst. of Technology, Atlanta, 2004).
- [31] Y. Lan and P. Cvitanović, “Variational method for finding periodic orbits in a general flow”, *Phys. Rev. E* **69**, 016217 (2004).
- [32] Y. Lan and P. Cvitanović, “Unstable recurrent patterns in Kuramoto-Sivashinsky dynamics”, *Phys. Rev. E* **78**, 026208 (2008).
- [33] R. E. LaQuey, S. M. Mahajan, P. H. Rutherford, and W. M. Tang, “Non-linear saturation of the trapped-ion mode”, *Phys. Rev. Lett.* **34**, 391–394 (1974).
- [34] D. Michelson, “Steady solutions of the Kuramoto-Sivashinsky equation”, *Physica D* **19**, 89–111 (1986).

- [35] A. A. Nepomnyashchii, “Stability of wavy conditions in a film flowing down an inclined plane”, *Fluid Dynamics* **9**, 354–359 (1975).
- [36] W. M. F. Orr, “The stability or instability of the steady motions of a liquid. Part II: A viscous liquid”, *Proc. R. Irish. Acad. A* **27**, 69–138 (1907).
- [37] O. Reynolds, “An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous, and the law of resistance in parallel channels”, *Proc. R. Soc. Lond. A* **174**, 935–982 (1883).
- [38] G. I. Sivashinsky, “Nonlinear analysis of hydrodynamical instability in laminar flames - I. Derivation of basic equations”, *Acta Astronaut.* **4**, 1177–1206 (1977).
- [39] A. P. Willis, *Equilibria, periodic orbits and computing them*, 2019.
- [40] Z. Wu, T. A. Zaki, and C. Meneveau, “High-Reynolds-number fractal signature of nascent turbulence during transition”, *Proc. Natl. Acad. Sci.* **117**, 3461–3468 (2020).
- [41] H.-l. Yang, K. A. Takeuchi, F. Ginelli, H. Chaté, and G. Radons, “Hyperbolicity and the effective dimension of spatially-extended dissipative systems”, *Phys. Rev. Lett.* **102**, 074102 (2009).

30.8 Examples

Example 30.1. Kuramoto-Sivashinsky antisymmetric subspace: The Fourier coefficients a_k are in general complex numbers. We can isolate the antisymmetric subspace $u(x, t) = -u(-x, t)$ by considering the case of a_k pure imaginary, $a_k \rightarrow ia_k$, where $a_k = -a_{-k}$ are real, with the evolution equations

$$\dot{a}_k = q_k^2 (1 - q_k^2) a_k + \frac{q_k}{2} \sum_{m=-\infty}^{+\infty} a_m a_{k-m}. \quad (30.38)$$

By picking this subspace we eliminate the continuous translational symmetry from our considerations; that is not an option for an experimentalist, but will do for our purposes. In the antisymmetric subspace the translational invariance of the full system reduces to the invariance under discrete translation by half a spatial period L . In the Fourier representation (30.38) this corresponds to invariance under

$$a_{2m} \rightarrow a_{2m}, a_{2m+1} \rightarrow -a_{2m+1}. \quad (30.39)$$

[return to p. 612](#)

Example 30.2. Cyclic subgroups of $SO(2)$: Any rational shift $\tau_{1/m} u(x) = u(x + L/m)$ generates a discrete cyclic subgroup C_m of $O(2)$, also a symmetry of KS system. Reflection together with C_m generates another symmetry of KS system, the dihedral subgroup D_m of $O(2)$. The only non-zero Fourier components of a solution invariant under C_m are $a_{jm} \neq 0$, $j = 1, 2, \dots$, while for a solution invariant under D_m we also have the condition $\text{Re } a_j = 0$ for all j . D_m reduces the dimensionality of state space and aids computation of equilibria and periodic orbits within it. For example, the 1/2-cell translations

$$\tau_{1/2} u(x) = u(x + L/2) \quad (30.40)$$

and reflections generate $O(2)$ subgroup $D_2 = \{1, \sigma, \tau, \tau\sigma\}$, which reduces the state space into four irreducible subspaces (for brevity, here $\tau = \tau_{1/2}$):

$$\begin{array}{c} \tau \quad \sigma \quad \tau\sigma \\ P^{(1)} = \frac{1}{4}(1 + \tau + \sigma + \tau\sigma) \quad S \quad S \quad S \\ P^{(2)} = \frac{1}{4}(1 + \tau - \sigma - \tau\sigma) \quad S \quad A \quad A \\ P^{(3)} = \frac{1}{4}(1 - \tau + \sigma - \tau\sigma) \quad A \quad S \quad A \\ P^{(4)} = \frac{1}{4}(1 - \tau - \sigma + \tau\sigma) \quad A \quad A \quad S. \end{array} \quad (30.41)$$

$P^{(j)}$ is the projection operator onto $u^{(j)}$ irreducible subspace, and the last 3 columns refer to the symmetry (or antisymmetry) of $u^{(j)}$ functions under reflection and 1/2-cell shift. By the same argument that identified (30.12) as the invariant subspace of KS, here the KS flow stays within the $\mathbb{U}^S = \mathbb{U}^{(1)} + \mathbb{U}^{(2)}$ irreducible D_1 subspace of u profiles symmetric under 1/2-cell shifts.

While in general the bilinear term $(u^2)_x$ mixes the irreducible subspaces of D_n , for D_2 there are four subspaces invariant under the flow [27]:

$\{0\}$: the $u(x) = 0$ equilibrium

$\mathbb{U}^+ = \mathbb{U}^{(1)} + \mathbb{U}^{(3)}$: the reflection D_1 irreducible space of antisymmetric $u(x)$

$\mathbb{U}^S = \mathbb{U}^{(1)} + \mathbb{U}^{(2)}$: the shift D_1 irreducible space of $L/2$ shift symmetric $u(x)$

$\mathbb{U}^{(1)}$: the D_2 irreducible space of $u(x)$ invariant under $x \mapsto L/2 - x$, $u \mapsto -u$.

With the continuous translational symmetry eliminated within each subspace, there are no relative equilibria and relative periodic orbits, and one can focus on the equilibria and periodic orbits only, as was done for $\mathbb{U}^+$ in refs. [9, 30, 32]. In the Fourier representation, the $u \in \mathbb{U}^+$ antisymmetry amounts to having purely imaginary coefficients, since $a_{-k} = a_k^* = -a_k$. The $1/2$ cell-size shift $\tau_{1/2}$ generated 2-element discrete subgroup $\{1, \tau_{1/2}\}$ is of particular interest because in the $\mathbb{U}^+$ subspace the translational invariance of the full system reduces to invariance under discrete translation (30.40) by half a spatial period $L/2$.

Each of the above dynamically invariant subspaces is unstable under small perturbations, and generic solutions of Kuramoto-Sivashinsky equation belong to the full space. Nevertheless, since all equilibria of the KS flow studied in this paper lie in the $\mathbb{U}^+$ subspace (see sect. 30.6), $\mathbb{U}^+$ plays important role for the global geometry of the flow. However, linear stability of these equilibria has eigenvectors both in and outside of $\mathbb{U}^+$, and needs to be computed in the full state space.

[return to p. 612](#)

Example 30.3. Kuramoto-Sivashinsky simulation, antisymmetric subspace: To get started, we set $\nu = 0.029910$, $L = 2\pi$ in the Kuramoto-Sivashinsky equation (30.2), or, equivalently, $\nu = 1$, $L = 36.33052$ in the non-dimensionalized units (30.38). Consider the antisymmetric subspace (30.38), so the non-dimensionalized system size is $\tilde{L} = L/2\pi = 2.89109$. Truncate (30.38) to $0 \leq k \leq 16$, and integrate an arbitrary initial condition. Let the transient behavior settle down.

Why this $\tilde{L}$? For this system size $\tilde{L}$ the dynamics appears to be chaotic, as far as can be determined numerically. Why $N = 16$? In practice one repeats the same calculation at different truncation cutoffs N , and makes sure that the inclusion of additional modes has no effect within the desired accuracy. For this system size $N = 16$ suffices.

Once a trajectory is computed in Fourier space, we can recover and plot the corresponding spatiotemporal pattern $u(x, t)$ over the configuration space using (2.19), as in figure 30.1 and figure 30.5. Such patterns give us a qualitative picture of the flow, but no detailed dynamical information; for that, tracking the evolution in a high-dimensional state space, such as the space of Fourier modes, is much more informative.

[return to p. 620](#)

Example 30.4. Kuramoto-Sivashinsky return maps: Consider the Kuramoto-Sivashinsky equation in the N Fourier modes representation. We pick (arbitrarily) the hyperplane $a_1 = 0$ as the Poincaré section, and integrate (30.14) with $a_1 = 0$, and an arbitrary initial point $(a_2, \dots, a_N)$. When the flow crosses the $a_1 = 0$ hyperplane in the same direction as initially, the initial point is mapped into $(a'_2, \dots, a'_N) = P(a_2, \dots, a_N)$. This defines P , the return map (3.1) of the $(N - 1)$ -dimensional $a_1 = 0$ hyperplane into itself.

Figure 30.7 is a typical result. We have picked - again arbitrarily - a subspace such as $a_6(n + 1)$ vs. $a_6(n)$ in order to visualize the dynamics. While the topology of the attractor is still obscure, one thing is clear: even though the flow state space is infinite dimensional, the attractor is finite and thin, barely thicker than a line.

[return to p. 612](#)

Example 30.5. Stability matrix: The Kuramoto-Sivashinsky flat flame front $u(x, t) = 0$ is an equilibrium point of (30.2). The stability matrix (4.3) follows from (30.14)

$$A_{kj}(a) = \frac{\partial v_k(a)}{\partial a_j} = (q_k^2 - q_k^4)\delta_{kj} + q_k(a_{k-j} - a_{k+j}). \quad (30.42)$$

For the $u(x, t) = 0$ equilibrium solution the stability matrix is diagonal, the eigenvectors are Fourier modes, and – as in (4.32) – the Jacobian matrix is diagonal, $J_{kj}^t(0) = \delta_{kj}e^{(q_k^2 - q_k^4)t}$.

Example 30.6. Equilibria of equilibria.

The equilibrium condition $u_t = 0$ for the Kuramoto-Sivashinsky equation PDE (30.2) is the ODE

$$\frac{1}{2}(u^2)_x + u_{xx} + u_{xxx} = 0.$$

More generally, the relative equilibrium condition (30.18) is the ODE

$$\frac{1}{2}(u^2)_x + u_{xx} + u_{xxx} = c u_x \quad (30.43)$$

which can be analyzed as a dynamical system in its own right. Integrating once we get

$$\frac{1}{2}u^2 - cu + u_x + u_{xxx} = E. \quad (30.44)$$

This equation can be interpreted as a 3-dimensional dynamical system with spatial coordinate x playing the role of ‘time,’ and the integration constant E can be interpreted as ‘energy,’ see sect. 30.3.

The value of E strongly influences the nature of the solutions. For $E > 0$ there is rich E -dependent dynamics, with fractal sets of bounded solutions investigated in depth by Michelson [34]. For $\tilde{L} < 1$ the only equilibrium of the system is the globally attracting constant solution $u(x, t) = 0$, denoted EQ_0 from now on. With increasing system size L the system undergoes a series of bifurcations. The resulting equilibria and relative equilibria are described in the classical papers of Kevrekidis, Nicolaenko and Scovel [27], and Greene and Kim [21], among others. The relevant bifurcations up to the system size investigated here are summarized in figure 30.2: at $\tilde{L} = 22/2\pi = 3.5014\dots$, the equilibria are the constant solution EQ_0 , the equilibrium EQ_1 called GLMRT by Greene and Kim [21, 33], the 2- and 3-cell states EQ_2 and EQ_3 , and the pairs of relative equilibria $TW_{\pm 1}$, $TW_{\pm 2}$. All equilibria are in the antisymmetric subspace $\mathbb{U}^+$, while EQ_2 is also invariant under D_2 and EQ_3 under D_3 .

Written as a 3-dimensional dynamical system (30.44) (considering only the equilibrium, $c = 0$ case for now) with spatial coordinate x playing the role of “time,” this is a volume preserving flow

$$u_x = v, \quad v_x = w, \quad w_x = u^2 - v - E, \quad (30.45)$$

with the “time” reversal symmetry,

$$x \rightarrow -x, \quad u \rightarrow -u, \quad v \rightarrow v, \quad w \rightarrow -w.$$

From (30.45) we see that

$$(u + w)_x = u^2 - E.$$

If $E < 0$, $u + w$ increases without bound with $x \rightarrow \infty$, and every solution escapes to infinity. If $E = 0$, the origin $(0, 0, 0)$ is the only bounded solution.

For $E > 0$ there is much E -dependent interesting dynamics, with complicated fractal sets of bounded solutions. The sets of the solutions of the equilibrium condition (30.45) are themselves in turn organized by the equilibria of the equilibrium condition, and the connections between them. For $E > 0$ the equilibrium points of (30.45) are $c_+ = (\sqrt{E}, 0, 0)$ and $c_- = (-\sqrt{E}, 0, 0)$. Linearization of the flow around c_+ yields Floquet exponents $[2\lambda, -\lambda \pm i\theta]$ with

$$\lambda = \frac{1}{\sqrt{3}} \sinh \phi, \quad \theta = \cosh \phi,$$

and ϕ fixed by $\sinh 3\phi = 3\sqrt{3E}$. Hence c_+ has a 1-dimensional unstable manifold and a 2-dimensional stable manifold along which solutions spiral in. By the $x \rightarrow -x$ “time reversal” symmetry, the invariant manifolds of c_- have reversed stability properties.

The non-wandering set of this dynamical system is quite pretty, and surprisingly hard to analyze. However, we do not need to explore the fractal set of the Kuramoto-Sivashinsky equilibria for infinite size system here; for a fixed system size L with periodic boundary condition, the only surviving equilibria are those with periodicity L . They satisfy the equilibrium condition for (30.14)

$$q_k^2 (1 - q_k^2) a_k - i \frac{q_k}{2} \sum_{m=-\infty}^{+\infty} a_m a_{k-m} = 0. \quad (30.46)$$

Periods of spatially periodic equilibria are multiples of L . Every time $\tilde{L}$ crosses an integer value $\tilde{L} = n$, new n -cell states are generated through pitchfork bifurcations. In the full state space they form an invariant circle due to the translational invariance of (30.4). In the antisymmetric subspace (see example 30.1), they corresponds to two points, half-period translates of each other of the form

$$u(x, t) = -2 \sum_k a_{kn} \sin(knx),$$

where $a_{kn} \in \mathbb{R}$.

For any fixed spatial period L the number of spatially periodic solutions is finite up to a spatial translation. This observation can be heuristically motivated as follows. Finite dimensionality of the inertial manifold bounds the size of Fourier components of all solutions. On a finite-dimensional compact manifold, an analytic function can only have a finite number of zeros. So, the equilibria, i.e., the zeros of a smooth velocity field on the inertial manifold, are finitely many.

For a sufficiently small L the number of equilibria is small, mostly concentrated on the low wave number end of the Fourier spectrum. These solutions may be obtained by solving the truncated versions of (30.46).

(Y. Lan and P. Cvitanović)

[return to p. 614](#)

Example 30.7. Some Kuramoto-Sivashinsky equilibria:

See figure 30.11.

See figure 30.12.

[return to p. 614](#)

Figure 30.11: Long-time evolution of a typical “sustained turbulence” trajectory for $L = 38.5$ in: (a) The space-time representation of $u(x, t)$ in the $[x, t]$ plane, $x \in [0, L/2]$ horizontally, $t \in [0, 7600]$ in vertical segments. The color represents the magnitude of $u(x, t)$. (b) $[a_1, a_2]$ Fourier modes projection, (c) $[a_3, a_4]$ projection. The typical time scale is set by the shortest periods of the periodic orbits embedded in the central, “wobbly” and side, “traveling wave” patterns of order $T = 20 \sim 25$, so this is a very long simulation, over 300 “turnover” times. The goal of this paper is to describe the characteristic unstable “wobble” and “traveling wave” patterns in terms of a hierarchy of invariant periodic orbit solutions (from ref. [32]).

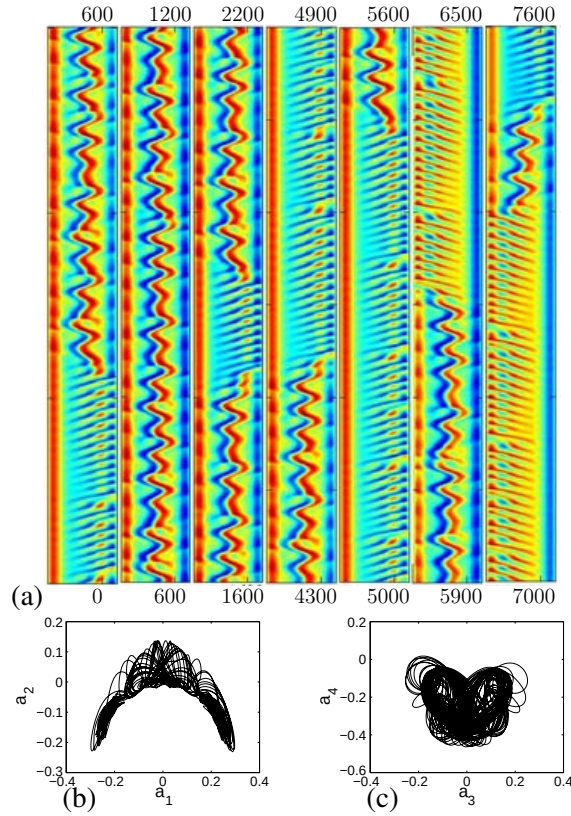
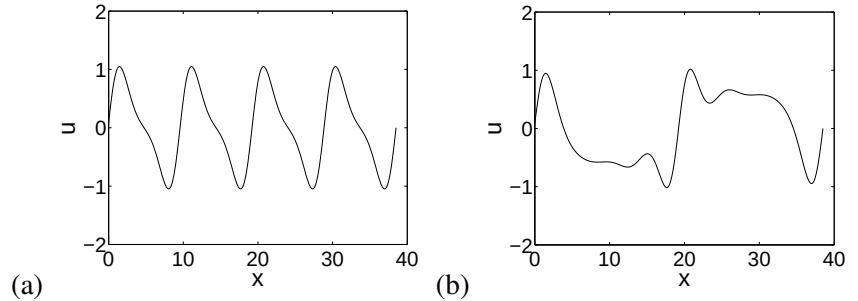


Figure 30.12: The non-wandering set displayed in figure 30.11 (b) and (c) appears to consist of three patches: the left part (S_L), the center part (S_C) and the right part (S_R), each centered around an unstable equilibrium: (a) central C_1 equilibrium, (b) right side R_1 equilibrium on the interval $[0, L]$, and its reflection L_1 (from ref. [32]).



30.9 Exercises

30.1. Galilean invariance of the Kuramoto-Sivashinsky equation.

- Verify that the Kuramoto-Sivashinsky equation is Galilean invariant: if $u(x, t)$ is a solution, then $v + u(x + 2vt, t)$, with v an arbitrary constant velocity, is also a solution.
- Verify that mean

$$\langle u \rangle = \frac{1}{L} \int_L dx u$$

is conserved by the flow.

- Argue that the choice (30.3) of the vanishing mean velocity, $\langle u \rangle = 0$ leads to no loss of generality in calculations that follow.

- (d)  [thinking is extra cost] Inspection of various “turbulent” solutions of Kuramoto-Sivashinsky equation reveals subregions of “traveling waves” with locally nonzero $\langle u \rangle$. Is there a way to use Galilean invariance locally, even though we eliminated it by the $\langle u \rangle = 0$ condition?

30.2. Infinite dimensional dynamical systems are not smooth. Many of the operations we consider natural for finite dimensional systems do not have smooth behavior in infinite dimensional vector spaces. Consider, as an example, a concentration ϕ diffusing on $\mathbb{R}$ according to the diffusion equation

$$\partial_t \phi = \frac{1}{2} \nabla^2 \phi.$$

- (a) Interpret the partial differential equation as an infinite dimensional dynamical system. That is, write it as $\dot{x} = F(x)$ and find the velocity field.
 (b) Show by examining the norm

$$\|\phi\|^2 = \int_{\mathbb{R}} dx \phi^2(x)$$

that the vector field F is not continuous.

- (c) Try the norm

$$\|\phi\| = \sup_{x \in \mathbb{R}} |\phi(x)|.$$

Is F continuous?

- (d) Argue that the semi-flow nature of the problem is not the cause of our difficulties.
 (e) Do you see a way of generalizing these results?

30.3. Kuramoto-Sivashinsky energy transfer rates.

- (a) Derive (30.30) from (30.29). Now that you have your integration by parts skills hone, also show that

$$\begin{aligned} \langle u_{xxx} u^2 \rangle &= \langle u_x^3 \rangle \\ \langle u_{xxxx} u^2 \rangle &= -5 \langle u_x u_{xx}^2 \rangle. \end{aligned} \quad (30.47)$$

- (b) Derive the power - dissipation rate relation (30.33).
 (c) Prove that for an equilibrium E is constant.
 (d) Derive formulas for $\dot{P}$, $\dot{D}$, $\dot{E}$ and $\frac{d}{dt} \langle u_x^3 \rangle$ in terms of space averages $\langle \cdots \rangle$. You will note that higher derivatives of u appear. The guiding principle is to use integration by parts until the number of such derivatives is minimized.
 (e) Invent another such formula.

30.4. Navier-Stokes energy transfer rates. The Millenium Prize tempts you to ponder the Navier-Stokes equations

$$\partial_t v_i + v_j \partial_j v_i = -\partial_i p + \nu \partial_{jj} v_i \quad (30.48)$$

in the utterly unphysical setting, a periodic 3D box of size $[L \times L \times L]$. The space average of a function $a = a(x, t) = a(v(x, t))$ on the interval L is given by

$$\langle a \rangle = \frac{1}{L^3} \oint dx^3 a(x, t). \quad (30.49)$$

- (a) Prove conservation of momentum

$$\frac{d}{dt} \langle v_i \rangle = 0 \quad (30.50)$$

- (b) Prove power-dissipation rate relation

$$\frac{1}{2} \frac{d}{dt} \langle v^2 \rangle = -\nu \langle |\omega|^2 \rangle \quad (30.51)$$

- (c) Prove conservation of helicity.

$$\frac{1}{2} \frac{d}{dt} \langle v \cdot \omega \rangle = -\nu \langle \omega \cdot \nabla \times \omega \rangle \quad (30.52)$$

- (d) While you are on the roll: derive another such formula. Pipe or plane Couette flow power-dissipation relation $\dot{E} = P - D$ would be particularly useful.

30.5. **Local Galilean invariance of Kuramoto-Sivashinsky?**



Inspection of various “turbulent” solutions of Kuramoto-Sivashinsky equation reveals subregions of “traveling waves” with locally nonzero $\langle u \rangle$. Is there a way to use Galilean invariance locally, even though we eliminated it by the $\langle u \rangle = 0$ condition?

Chapter 31

Koopman modes

(S. Bagheri and P. Cvitanović)

SO FAR WE HAVE mostly focused on computation of eigenvalues of evolution operators. Here we shall discuss the role of their eigenfunctions. This is easiest to explain for systems with stable equilibria and periodic orbits, for which the dynamics is described by Koopman operators. We shall show how here how the *nonlinear* dynamics of transient states on the way to a stable solution is captured by the eigenfunctions of the *linear* Koopman operator.



31.1 Koopmania

The Koopman operator action on an observable $a(x)$ (a bounded and smooth state space function that associates a scalar to state x) is to replace it by its downstream value time t later, $a(x) \rightarrow a(x(t))$, evaluated at the trajectory point $x(t)$:

$$\begin{aligned} [\mathcal{K}^t a](x) &= a(f^t(x)) = \int_{\mathcal{M}} dy \mathcal{K}^t(x, y) a(y) \\ \mathcal{K}^t(x, y) &= \delta(y - f^t(x)) . \end{aligned} \tag{31.1}$$

Given an initial density of representative points $\rho(x)$, the state space average of $a(x)$ evolves as

$$\begin{aligned} \langle a \rangle_{\rho(t)} &= \frac{1}{|\rho_{\mathcal{M}}|} \int_{\mathcal{M}} dx a(f^t(x)) \rho(x) = \frac{1}{|\rho_{\mathcal{M}}|} \int_{\mathcal{M}} dx [\mathcal{K}^t a](x) \rho(x) \\ &= \frac{1}{|\rho_{\mathcal{M}}|} \int_{\mathcal{M}} dx dy a(y) \delta(y - f^t(x)) \rho(x) . \end{aligned}$$

The ‘propagator’ $\delta(y - f^t(x))$ can be interpreted as belonging to the Perron-Frobenius operator (19.10), so the two operators are adjoint to each other,

$$\int_{\mathcal{M}} dx [\mathcal{K}^t a](x) \rho(x) = \int_{\mathcal{M}} dy a(y) [\mathcal{L}^t \rho](y). \quad (31.2)$$

The Koopman and Perron-Frobenius operators describe the dynamics in complementary ways. Koopman advances the trajectory by time t , Perron-Frobenius depends on the trajectory point time t in the past. Perron-Frobenius propagates a conserved quantity (a density of initial conditions) forward in time. The growth (or decay) of the density depends on the compression (or expansion) of a volume occupied by a set of trajectories. The dynamics of an observable depends on the other hand on one single trajectory.

exercise 31.1

The family of Koopman operators $\{\mathcal{K}^t\}_{t \in \mathbb{R}_+}$ forms a semigroup parameterized by time, $\mathcal{K}^t \mathcal{K}^{t'} = \mathcal{K}^{t+t'}$, $\mathcal{K}^0 = \mathbf{1}$ with the generator of infinitesimal time translations defined by

$$\mathcal{A}^\dagger = \lim_{t \rightarrow 0^+} \frac{1}{t} (\mathcal{K}^t - \mathbf{1}).$$

If the flow is finite-dimensional and invertible, $\mathcal{A}^\dagger$ is a generator of a group. The explicit form of $\mathcal{A}^\dagger$ follows from expanding dynamical evolution up to first order, as in (2.8):

$$\mathcal{A}^\dagger a(x) = \lim_{t \rightarrow 0^+} \frac{1}{t} (a(f^t(x)) - a(x)) = v_i(x) \partial_i a(x). \quad (31.3)$$

This is by definition the time derivative, so the time-evolution equation for $a(x)$ is

$$\left(\frac{d}{dt} - \mathcal{A}^\dagger \right) a(x) = 0. \quad (31.4)$$

We formally write the solution to (31.4) as

appendix A31.2

$$a(x(t)) = e^{t\mathcal{A}^\dagger} a(x_0) = \mathcal{K}^t a(x_0),$$

so the finite time Koopman operator (31.1) can be recovered by exponentiating the time-evolution generator $\mathcal{A}^\dagger$. The generator $\mathcal{A}^\dagger$ looks very much like the generator of translations. For example, for a constant velocity field dynamical evolution is nothing but a translation by time $\times$ velocity:

exercise A31.1

exercise 19.10

$$e^{tv \frac{\partial}{\partial x}} a(x) = a(x + tv). \quad (31.5)$$

As we will not need to implement a computational formula for general $e^{t\mathcal{A}^\dagger}$ in what follows, we relegate making sense of such operators to appendix A31.2.

appendix A31.2

The Koopman / Perron-Frobenius operators are non-normal, non-self-adjoint operators, so their left and right eigenvectors differ. The right eigenvectors of a Perron-Frobenius operator are the left eigenvectors of the Koopman, and vice versa. That is,

$$\mathcal{A} \phi_\alpha(x) = s_\alpha \phi_\alpha(x), \quad \mathcal{A}^\dagger \psi_\alpha(x) = s_\alpha^* \psi_\alpha(x), \quad \alpha = 0, 1, 2, \dots$$

The left and right eigenfunctions satisfy the bi-orthogonality condition with respect to L^2 norm,

$$\int_{\mathcal{M}} dx \phi_{\alpha}^* \psi_{\beta} = \delta_{\alpha\beta}. \quad (31.6)$$

While one might think of a Koopman operator as an ‘inverse’ of the Perron-Frobenius operator, the notion of *adjoint* is the right one, especially in settings where flow is not time-reversible, as is the case for dissipative PDEs (infinite dimensional flows contracting forward in time) and for stochastic flows.

Given the left and right eigenfunctions, we can express the evolution of an observable as

$$a(x(t)) = [\mathcal{K}^t a](x_0) = \sum_{\alpha} c_{\alpha} e^{s_{\alpha} t} \psi_{\alpha}(x_0) \quad (31.7)$$

where

$$c_{\alpha} = \int_{\mathcal{M}} dx a(x) \phi_{\alpha}^*(x).$$

This expansion suggests an alternative description of nonlinear dynamics, which is the (linear) evolution of observables in an infinite-dimensional space. In principle, this allows the study of full nonlinear dynamics using linear operator-theoretical tools.



example 31.1
p. 645



example 31.2
p. 645

31.2 Koopman eigenvalues for a limit cycle

The $[(d-1) \times (d-1)]$ -dimensional monodromy matrix $\mathbf{M}_{ij} = \partial_j P_i(\hat{x}_a)$ of dimension governs the dynamics of the small perturbation $\delta \hat{x}$ within a Poincaré section.

Even though the monodromy matrix $\mathbf{M}(\hat{x})$ depends upon $\hat{x}$ (the ‘starting’ point of the periodic orbit), its eigenvalues do not, so we may write for its eigenvectors $\mathbf{e}^{(j)}$ (sometimes referred to as ‘covariant Lyapunov vectors,’ or, for periodic orbits, as ‘Floquet vectors’)

$$\mathbf{M}(x) \mathbf{e}^{(j)}(x) = \Lambda_j \mathbf{e}^{(j)}(x), \quad \Lambda_j = e^{\lambda^{(j)} T}. \quad (31.8)$$

where Floquet exponents $\lambda^{(j)} = \mu^{(j)} \pm i\omega^{(j)}$ are independent of x . We order the Floquet multipliers as

$$|\Lambda_1| \geq |\Lambda_2| \geq \dots \geq |\Lambda_{d-1}|. \quad (31.9)$$

The limit cycle is stable if $|\Lambda_1| < 1$.

The two most important characteristics of the limit cycle are thus the fundamental frequency and the leading Lyapunov exponent, defined by

$$\omega = \frac{2\pi}{T}, \quad \mu = \frac{1}{T} \ln |\Lambda_1|, \quad (31.10)$$

respectively.

Here we follow the derivations chapter 21, except that the analysis is restricted to the simpler case of a stable limit cycle. The trace of the Koopman operator is,

$$\text{tr } \mathcal{K}^t = \int_{\mathcal{M}} \mathcal{K}^t(x, x) dx.$$

where $\mathcal{K}^t$ is the kernel. Inspired by this definition, we define the trace of Koopman operator as

$$\text{tr } \mathcal{K}^t = \int_{\mathcal{M}} \delta(x - f^t(x)) dx. \quad (31.11)$$

From (31.11), one observes that the trace $\mathcal{K}^t$ receives a contribution whenever the trajectory returns to the starting point after r repeats of the limit cycle period T .

To proceed, we decompose the propagator f_t into two parts, the $(d-1)$ -dimensional return map P and a 1-dimensional return-time function τ . The return map captures only the transverse part of the periodic dynamics, since the flow component tangent to the trajectory, which is not in the span of the Poincaré section, has not been taken into account. Assuming the longitudinal state component has a certain mean velocity v as it traverses the limit cycle, one may transform this component to a time coordinate system using the relation $v dt$. Thus the full dynamics is described by the return map P and by the first return function $\tau(\hat{x})$ that provides the (non-constant) time interval between successive points $\hat{x}$ on Poincaré section, e.g. $t_{k+1} = t_k + \tau(\hat{x}_k)$. Applying τ recursively, we may write $((k+1)$ th time as a function first point and initial time,

$$t_{k+1} = t_1 + \sum_{j=0}^{k-1} \tau(P^j \hat{x}_1). \quad (31.12)$$

Now, factor the kernel of $\mathcal{K}^t$ (31.11) into two parts

$$\text{tr } \mathcal{K}^t = \int_{\mathcal{P}(\hat{x})=0} d\hat{x} \int_0^{\tau(\hat{x})} dt \delta(\hat{x} - P^k \hat{x}) \delta\left(t - \sum_{j=0}^{k-1} \tau(P^j \hat{x})\right), \quad (31.13)$$

where P^k and τ are defined above and in (31.12), respectively. We treat the two Dirac delta functions separately, starting with P^k . First recall that the Dirac delta function applied to a scalar-valued function $g(x)$, is

$$\int \delta(g(x)) dx = \int \delta(x) |g'(0)|^{-1} dx = \sum_j \frac{1}{|g'(x_j)|},$$

where x_j are the roots of $g(x)$. This property may be generalized to $d-1$ dimensions and applied to the Dirac-delta in (31.13),

$$\int_{\mathcal{P}(\mathbf{u})=0} d\hat{x} \delta(\hat{x} - P^k(\hat{x})) = \frac{1}{|\det(\mathbf{I} - \mathbf{M}^r)|}, \quad (31.14)$$

where $\mathbf{I}$ denotes the identity matrix. The second part of the trace can be written as

$$\int_0^{\tau(\hat{x})} \delta(t - \sum_{j=0}^{k-1} \tau(P^j \hat{x})) dt = T \sum_{r=1}^{\infty} \delta(t - rT). \quad (31.15)$$

Inserting the identities (31.14) and (31.15) in (31.13), we get the trace formula for a single limit cycle of period T ,

$$\text{tr } \mathcal{K}^t = T \sum_{r=1}^{\infty} \frac{\delta(t - rT)}{|\det(\mathbf{I} - \mathbf{M}^r)|}, \quad (31.16)$$

which was first derived in ref. [1], here given in the special case of a single limit cycle. The trace formula is a sum whose terms are nonzero only for integers of the cycle period. The r th nonzero term describes how much after the r th return to the Poincaré section a small neighborhood volume (i.e. a tube) of the stable limit cycle has retracted. This relation thus connects the trace of $\mathcal{K}^t$ to the dynamics in the local stable manifold of the limit cycle.

The Koopman eigenvalues are the poles of the Laplace transform of trace of $\mathcal{K}^t$

$$\int_0^{\infty} e^{-st} \text{tr } \mathcal{K}^t dt = \text{tr } \frac{1}{s - \mathcal{A}},$$

i.e., the poles of the resolvent of $\mathcal{A}$. By inserting (31.16) in the left-hand side of above equation one obtains,

$$\text{tr } \frac{1}{s - \mathcal{A}} = \frac{\partial}{\partial s} \ln(\det(s - \mathcal{A})),$$

where $\det(s - \mathcal{A})$ is the spectral determinant,

$$\det(s - \mathcal{A}) = \exp \left[- \sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{-sTr}}{|\det(\mathbf{I} - \mathbf{M}^r)|} \right].$$

Now, since the determinant does not depend on the basis which $\mathbf{M}$ is described in, we may write it in terms of the eigenvalues of $\mathbf{M}$,

$$\frac{1}{|\det(\mathbf{I} - \mathbf{M}^r)|} = \prod_{k=1}^{d-1} \frac{1}{1 - \Lambda_k^r}, \quad (31.17)$$

where we have assumed that $|\Lambda_k| < 1$ for all k .

Denominators can be expanded in Taylor series such as

$$(1 - x)^{-1}(1 - y)^{-1} = 1 + x + y + x^2 + xy + y^2 + \dots,$$

when $|x| < 1, |y| < 1$. Each term in the product (31.17) may thus be written as an infinite sum. Define a *multi-index* as an array of d non-negative integers $j_k = 0, 1, 2, \dots$:

$$\mathbf{j} = [j_1, j_2, \dots, j_d] \in \mathbb{N}^d,$$

Consider next the product of $d - 1$ Floquet multipliers

$$\Lambda = \Lambda_1 \Lambda_2 \cdots \Lambda_{d-1} = e^{T(\mu^{(1)} + \mu^{(2)} + \cdots + \mu^{(d-1)})},$$

(the imaginary parts of complex pairs cancel in the exponent), and define

$$\boldsymbol{\mu} = [\mu^{(1)}, \mu^{(2)}, \dots, \mu^{(d-1)}] \in \mathbb{R}^d.$$

Λ can now be raised to $\mathbf{j}$ th power as

$$\Lambda^{\mathbf{j}} = e^{T\boldsymbol{\mu} \cdot \mathbf{j}} = \Lambda_1^{j_1} \Lambda_2^{j_2} \cdots \Lambda_{d-1}^{j_{d-1}}. \quad (31.18)$$

Using multi-index notation (31.18) we may write (31.17) as

$$\frac{1}{|\det(\mathbf{I} - \mathbf{M}^r)|} = \sum_{\mathbf{j}} \Lambda^{\mathbf{j}r},$$

and consequently the spectral determinant as

$$\det(s - \mathcal{A}) = \exp \left[- \sum_{r=1}^{\infty} \frac{1}{r} (e^{-sT} \sum_{\mathbf{j}} \Lambda^{\mathbf{j}})^r \right].$$

Applying the identity $\sum x^r/r = -\ln(1-x)$, we obtain the final form of the spectral determinant for a stable limit cycle

$$\det(s - \mathcal{A}) = \prod_{\mathbf{j}} \left(1 - e^{-sT} \Lambda^{\mathbf{j}} \right). \quad (31.19)$$

The zeros of $\det(s - \mathcal{A}) = 0$ are given by the zeros of individual terms in the product:

$$e^{-T(s - \boldsymbol{\mu} \cdot \mathbf{j})} = 1.$$

Taking the logarithm of both sides, we obtain

$$s_{j,m} = \boldsymbol{\mu} \cdot \mathbf{j} + 2\pi i m / T = \boldsymbol{\mu} \cdot \mathbf{j} + i m \omega \quad (31.20)$$

with $m = 0, \pm 1, \pm 2, \dots$. For our particular choice of analytic observables the spectrum of $\mathcal{K}^t$ is reduced to its minimal components, namely any integer multiple of the stability eigenvalues. Thus, for any stable limit cycle, the Koopman eigenvalues form a lattice on the lower half of the complex plane. The marginal eigenvalues on the horizontal imaginary axis corresponding to $j = 0$ correspond to the non-decaying time-averaged mean ($m = 0$) and periodic dynamics ($m \neq 0$) on the limit cycle. The remaining eigenvalues $j \neq 0$ are decaying and describe the transient behavior of flow in the local stable manifold of the limit cycle.



example 31.3
p. 646

Commentary

Remark 31.1. Koopman operators. The “Heisenberg picture” in dynamical systems theory has been introduced by Koopman and Von Neumann [3, 5], see also ref. [4]. Inspired by the contemporary advances in quantum mechanics, Koopman [3] observed in 1931 that $\mathcal{K}^t$ is unitary on $L^2(\mu)$ Hilbert spaces. The Koopman operator is the classical analogue of the quantum evolution operator $\exp(i\hat{H}t/\hbar)$ – the kernel of $\mathcal{L}^t(y, x)$ introduced in (19.14) (see also sect. 20.2) is the analogue of the Green’s function discussed here in chapter 36. The relation between the spectrum of the Koopman operator and deterministic ergodicity was formalized by von Neumann [5]. We shall not use Hilbert spaces here and the operators that we shall study *will not* be unitary. For a discussion of the relation between the Perron-Frobenius operators and the Koopman operators for finite dimensional deterministic invertible flows, infinite dimensional contracting flows, and stochastic flows, see Lasota-Mackey [4] and Gaspard [2].

References

- [1] P. Cvitanović and B. Eckhardt, “Periodic orbit expansions for classical smooth flows”, *J. Phys. A* **24**, L237–L241 (1991).
- [2] P. Gaspard, *Chaos, Scattering and Statistical Mechanics* (Cambridge Univ. Press, Cambridge UK, 1998).
- [3] B. O. Koopman, “Hamiltonian systems and transformations in Hilbert space”, *Proc. Natl. Acad. Sci. USA* **17**, 315 (1931).
- [4] A. Lasota and M. MacKey, *Chaos, Fractals, and Noise; Stochastic Aspects of Dynamics* (Springer, New York, 1994).
- [5] J. von Neumann, “Zusätze zur Arbeit “Zur Operatorenmethode in der klassischen Mechanik”. (German) [Additions to the work “On operator methods in classical mechanics”]”, *Ann. Math.* **33**, 789–791 (1932).

31.3 Examples

Example 31.1. Spectrum of a 1D linear system. Consider a 1D system with a single equilibrium

$$\dot{x} = \lambda x, \quad (31.21)$$

If the observable $a(x)$ is a smooth, real-analytical function, the Koopman operator spectrum can be identified from its Taylor expansion,

$$\begin{cases} s_k &= k\lambda \\ \phi_k &= \delta^{(k)}(x) \\ \psi_k &= x^k \end{cases} \quad \text{when } \lambda < 0 \quad (\text{attractor}) \quad (31.22)$$

and

$$\begin{cases} s_k &= -(k+1)\lambda \\ \phi_k &= x^k \\ \psi_k &= \delta^{(k)}(x) \end{cases} \quad \text{when } \lambda > 0 \quad (\text{repeller}) \quad (31.23)$$

for $k = 0, 1, \dots$. Here the superscript $^{(k)}$ refers to the k th derivative. We observe the duality between the right/left eigenfunctions and the repelling/attracting points. When $\lambda < 0$, any neighborhood of representative points shrinks to a point and asymptotically the density becomes a singular function. On the other hand, any smooth observable has the asymptotic limit $a(0)$. Koopman operator $\mathcal{K}^t$ is thus the appropriate evolution operator to represent the dynamics in stable manifolds, since the observable dynamics goes along with the flow.

[return to p. 640](#)

Example 31.2. Spectrum of a 1D nonlinear system. As an example of how the effects of nonlinearity are captured by expansion into eigenfunctions of the Koopman operator, consider the stable nonlinear system:

$$\dot{x} = \lambda x - x^3, \quad \lambda < 0 \quad (31.24)$$

where the only equilibrium point is the attracting fixed point $x_q = 0$. The difference between (31.24) and the linear system in (31.21), is the presence of a cubic nonlinear term. However, the nonlinear coordinate transformation

$$y = g(x) = \frac{x}{\sqrt{x^2 - \lambda}} \quad (31.25)$$

transforms (31.24) into a linear system $\dot{y} = \lambda y$, whose spectrum is already determined by (31.22). The Koopman spectrum in terms of the coordinate x is thus

$$\begin{cases} s_k &= k\lambda \\ \phi_k(x) &= \delta^{(k)}(x - g^{-1}(y))|_{y=0} \\ \psi_k(x) &= \left(x / \sqrt{x^2 - \lambda}\right)^k \end{cases} \quad (31.26)$$

where $k = 0, 1, \dots$ and the derivative of δ is with respect to y . Comparing to (31.22), the Koopman eigenvalues are not modified by the cubic nonlinear term in (31.24), but the term $\sqrt{x^2 - \lambda}$ appears in the Koopman eigenfunctions.

Consider the expansion (31.7) of a position $x(t)$ at time t considered as an observable, $a(x(t)) = x(t)$,

$$x(t) = \left(\frac{-\lambda}{x_0^2 - \lambda}\right)^{1/2} x_0 e^{\lambda t} + \frac{1}{\sqrt{-\lambda}} \left(\frac{x_0}{x_0^2 - \lambda}\right)^{3/2} e^{3\lambda t} + \dots \quad (31.27)$$

Figure 31.1: (black line) The trajectory $x(t)$ of (31.24) plotted on logarithmic scale as a function of time, for $\lambda = -0.6$. (red lines) Reconstructions of the trajectory based on the expansion (31.27) – including up to the ϕ_1, ϕ_3, ϕ_5 or ϕ_7 left eigenfunction of $\mathcal{K}^t$. (dashed line) The trajectory of the linearized system, with x^3 neglected in (31.24).

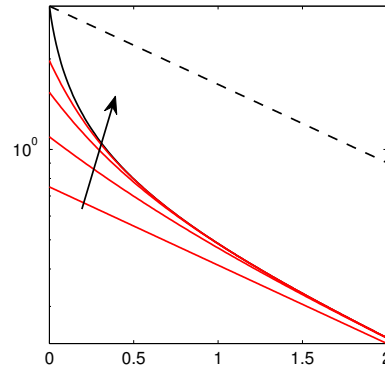
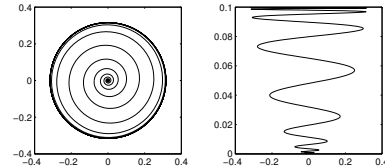


Figure 31.2: State trajectory starting close to $x_q = 0$ and with $\mu = 1/10$ for the system (31.28) in x, y -plane (left) and the (x, z) -plane (right).



In figure 31.1, the trajectory $x(t)$ (black line) obtained by integrating (31.24) starts out by a rapid decay to the stable manifold of the stable fixed point, followed by an exponential decay along the manifold to $x_q = 0$. In a purely linear analysis, the state evolves as $x_{lim}(t) = x_0 e^{\lambda t}$ (dashed black line in the figure). A linear analysis provides the exponential decay rate, but fails to describe the curved trajectory in its initial stages. In the figure the first non-zero expansion terms and the superposition of gradually increasing number of modes are shown with red lines. Whereas the Koopman eigenvalues provide the asymptotic decay rate, the Koopman eigenfunctions provide the direction as well as an amplitude. Including higher order terms in the expansion, eventually the full state trajectory can be recovered by a number of Koopman eigenfunctions, and thus the transient nonlinear dynamics preceding the infinitesimal linear region can be captured.

[return to p. 640](#)

Example 31.3. Spectrum of a stable limit cycle. Consider the three dimensional system

$$\begin{cases} \dot{x} &= \mu x - y - xz \\ \dot{y} &= \mu y - x - yz \\ \dot{z} &= -z + x^2 + y^2, \end{cases} \quad (31.28)$$

for $\mu \gtrsim 0$. The system has an unstable fixed point

$$x_q = (x, y, z) = 0,$$

and an attracting limit cycle

$$x_a = (\sqrt{\mu} \cos t, \sqrt{\mu} \sin t, \mu).$$

In figure 31.2, a typical trajectory starting near x_q is shown. The trajectory grows exponentially with the exponent $\lambda_q > 0$ and after a transient time, approaches the stable limit cycle exponentially fast with the exponent $\lambda_a < 0$.

The set of discrete Koopman/Perron-Frobenius eigenvalues is simply the union of the eigenvalues associated with the fixed point and the limit cycle. One may thus treat the two critical elements separately using the formulas derived from the trace of the operators.

Here we only consider the spectrum pertaining to stable limit cycle. By considering the Poincaré section given by the plane $y = 0$ and its associated monodromy matrix, one arrives at

$$\Lambda = -2\mu, \quad \omega = 2\pi.$$

According to formula (31.20) the Koopman/Perron-Frobenius eigenvalues $\{j, m\} = \{j_1, 0, \dots, 0, m\}$ corresponding to this leading Floquet exponent are,

$$s_{j,m} = j\Lambda + im\omega = -2\mu j + mi2\pi$$

for $j = 0, 1, 2, \dots$ and $m = 0, \pm 1, \pm 2, \dots$. The expansion of the state observable into the leading complex Koopman eigenfunctions ($j = 0, 1$ and $m = 0, 1$) associated with (31.28) is

$$x(t) = v_{0,0} + v_{0,1} e^{it} + v_{1,0} e^{-2\mu t} + c.c + \dots$$

with,

$$v_{0,0} = (0, 0, \mu), \quad (31.29)$$

$$v_{0,1} = \frac{\sqrt{\mu}}{2}(1, 0, 0) + \frac{i\sqrt{\mu}}{2}(0, 1, 0), \quad (31.30)$$

$$v_{1,0} = \frac{c\sqrt{\mu}}{2}(0, 0, \frac{r^2 - \mu}{r^2}), \quad (31.31)$$

where c is some constant and $r^2 = x^2 + y^2$.

The first two modes resolve the attractor dynamics; $v_{0,0}$ represents the average asymptotic value, and $v_{0,1}$ the periodic asymptotic solution with unit frequency on the attractor. These two Koopman modes correspond to the three first (real) empirical Karhunen-Loève or proper orthogonal decomposition modes. A robust low-order representation of the flow should in addition to the limit cycle also, at least in some sense, capture the dynamics of the corresponding attracting inertial manifold, that connects the unstable fixed point with the limit cycle. This is the role of the transient mode $v_{1,0}$; the function $(r^2 - \mu)/r^2$ is singular near the fixed point and zero at the limit cycle and points in the direction z , i.e. from x_q to x_a .

[return to p. 643](#)

Exercises

31.1. **Perron-Frobenius operator is the adjoint of the Koopman operator.** Check (31.2) - it might be wrong as it stands. Pay attention to presence/absence of a Jacobian.

31.2. **Nonlinear system mapped into a linear one.** (31.24) and the linear system in (31.21), is the presence of a cubic nonlinear term. Show the nonlinear coordinate transformation (31.25)

$$y = g(x) = \frac{x}{\sqrt{x^2 - \lambda}}$$

transforms (31.24) into a linear system $\dot{y} = \lambda y$.

31.3. **Stability of a limit cycle.** Show that the system (31.28) has an attracting limit cycle $x_a = (\sqrt{\mu} \cos t, \sqrt{\mu} \sin t, \mu)$.

Chapter 32

Irrationally winding

I don't care for islands, especially very small ones.

—D.H. Lawrence

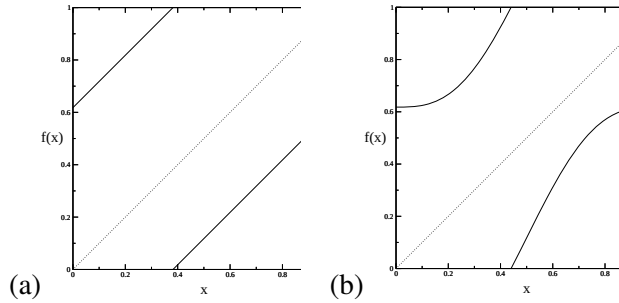
(R. Artuso and P. Cvitanović)

THIS CHAPTER is concerned with the mode locking problems for circle maps: besides its physical relevance it nicely illustrates the use of cycle expansions away from the dynamical setting, in the realm of renormalization theory at the transition to chaos.

The physical significance of circle maps is connected with their ability to model the two-frequencies mode-locking route to chaos for dissipative systems. In the context of *dissipative* dynamical systems one of the most common and experimentally well explored routes to chaos is the two-frequency mode-locking route. Interaction of pairs of frequencies is of deep theoretical interest due to the generality of this phenomenon; as the energy input into a dissipative dynamical system (for example, a Couette flow) is increased, typically first one and then two of intrinsic modes of the system are excited. After two Hopf bifurcations (a fixed point with inward spiralling stability has become unstable and outward spirals to a limit cycle) a system lives on a two-torus. Such systems tend to mode-lock: the system adjusts its internal frequencies slightly so that they fall in step and minimize the internal dissipation. In such case the ratio of the two frequencies is a rational number. An irrational frequency ratio corresponds to a quasiperiodic motion - a curve that never quite repeats itself. If the mode-locked states overlap, chaos sets in. The likelihood that a mode-locking occurs depends on the strength of the coupling of the two frequencies.

Our main concern in this chapter is to illustrate the “global” theory of circle maps, connected with universality properties of the whole irrational winding set. We shall see that critical global properties may be expressed via cycle expansions involving “local” renormalization critical exponents. The renormalization theory of critical circle maps demands rather tedious numerical computations, and our

Figure 32.1: (a) Unperturbed circle map ($k = 0$ in (32.1)) with golden mean rotation number. (b) Critical circle map ($k = 1$ in (32.1)) with golden mean bare rotation number.



intuition is much facilitated by approximating circle maps by number-theoretic models. The models that arise in this way are by no means mathematically trivial, they turn out to be related to number-theoretic abysses such as the Riemann conjecture, already in the context of the “trivial” models.

32.1 Mode locking

The simplest way of modeling a nonlinearly perturbed rotation on a circle is by 1-dimensional circle maps $x \rightarrow x' = f(x)$, restricted to the one dimensional torus, such as the *sine map*

$$x_{n+1} = f(x_n) = x_n + \Omega - \frac{k}{2\pi} \sin(2\pi x_n) \mod 1. \quad (32.1)$$

$f(x)$ is assumed to be continuous, have a continuous first derivative, and a continuous second derivative at the inflection point (where the second derivative vanishes). For the generic, physically relevant case (the only one considered here) the inflection is cubic. Here k parametrizes the strength of the nonlinear interaction, and Ω is the *bare* frequency.

The state space of this map, the unit interval, can be thought of as the elementary cell of the map

$$\hat{x}_{n+1} = \hat{f}(\hat{x}_n) = \hat{x}_n + \Omega - \frac{k}{2\pi} \sin(2\pi \hat{x}_n). \quad (32.2)$$

where $\hat{\cdot}$ is used in the same sense as in chapter 24.

The winding number is defined as

$$W(k, \Omega) = \lim_{n \rightarrow \infty} (\hat{x}_n - \hat{x}_0)/n. \quad (32.3)$$

and can be shown to be independent of the initial value $\hat{x}_0$.

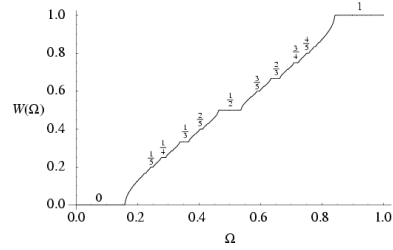
For $k = 0$, the map is a simple rotation (the *shift map*) see figure 32.1 (a)

$$x_{n+1} = x_n + \Omega \mod 1, \quad (32.4)$$

and the rotation number is given by the parameter Ω .

$$W(k = 0, \Omega) = \Omega.$$

Figure 32.2: The critical circle map ($k = 1$ in (32.1)), often referred to as the devil's staircase [21]; here the winding number W is a function of the parameter Ω .



For given values of Ω and k the winding number can be either rational or irrational. For invertible maps and rational winding numbers $W = P/Q$ the asymptotic iterates of the map converge to a unique attractor, a stable periodic orbit of period Q

$$\hat{f}^Q(\hat{x}_i) = \hat{x}_i + P, \quad i = 0, 1, 2, \dots, Q-1.$$

This is a consequence of the independence of $\hat{x}_0$ previously mentioned. There is also an unstable cycle, repelling the trajectory. For any rational winding number, there is a finite interval of values of Ω values for which the iterates of the circle map are attracted to the P/Q cycle. This interval is called the P/Q mode-locked (or stability) interval, and its width is given by

exercise 32.1

$$\Delta_{P/Q} = Q^{-2\mu_{P/Q}} = \Omega_{P/Q}^{right} - \Omega_{P/Q}^{left}. \quad (32.5)$$

where $\Omega_{P/Q}^{right}$ ($\Omega_{P/Q}^{left}$) denote the biggest (smallest) value of Ω for which $W(k, \Omega) = P/Q$. Parametrizing mode lockings by the exponent μ rather than the width Δ will be convenient for description of the distribution of the mode-locking widths, as the exponents μ turn out to be of bounded variation. The stability of the P/Q cycle is

$$\Lambda_{P/Q} = \frac{\partial x_Q}{\partial x_0} = f'(x_0)f'(x_1) \cdots f'(x_{Q-1})$$

For a stable cycle $|\Lambda_{P/Q}|$ lies between 0 (the superstable value, the “center” of the stability interval) and 1 (the $\Omega_{P/Q}^{right}$, $\Omega_{P/Q}^{left}$ endpoints of (32.5)). For the shift map (32.4), the stability intervals are shrunk to points. As Ω is varied from 0 to 1, the iterates of a circle map either mode-lock, with the winding number given by a rational number $P/Q \in (0, 1)$, or do not mode-lock, in which case the winding number is irrational. A plot of the winding number W as a function of the shift parameter Ω is a convenient visualization of the mode-locking structure of circle maps. It yields a monotonic “devil’s staircase” of figure 32.2 whose self-similar structure we are to unravel. Circle maps with zero slope at the inflection point x_c (see gmrot (b))

$$f'(x_c) = 0, \quad f''(x_c) = 0$$

($k = 1$, $x_c = 0$ in (32.1)) are called *critical*: they delineate the borderline of chaos in this scenario. As the nonlinearity parameter k increases, the mode-locked intervals become wider, and for the critical circle maps ($k = 1$) they fill out the whole interval. A critical map has a superstable P/Q cycle for any rational P/Q , as the stability of any cycle that includes the inflection point equals zero. If the

map is non-invertible ($k > 1$), it is called supercritical; the bifurcation structure of this regime is extremely rich and beyond the scope of this exposition.

The physically relevant transition to chaos is connected with the critical case, however the apparently simple “free” shift map limit is quite instructive: in essence it involves the problem of ordering rationals embedded in the unit interval on a hierarchical structure. From a physical point of view, the main problem is to identify a (number-theoretically) consistent hierarchy susceptible of experimental verification. We will now describe a few ways of organizing rationals along the unit interval: each has its own advantages as well as its drawbacks, when analyzed from both mathematical and physical perspective.

32.1.1 Hierarchical partitions of the rationals

Intuitively, the longer the cycle, the finer the tuning of the parameter Ω required to attain it; given finite time and resolution, we expect to be able to resolve cycles up to some maximal period Q . This is the physical motivation for partitioning mode lockings into sets of cycle period up to Q . In number theory such sets of rationals are called *Farey series*. They are denoted by $\mathcal{F}_Q$ and defined as follows. The Farey series of order Q is the monotonically increasing sequence of all irreducible rationals between 0 and 1 whose denominators do not exceed Q . Thus P_i/Q_i belongs to $\mathcal{F}_Q$ if $0 < P_i \leq Q_i \leq Q$ and $(P_i, Q_i) = 1$. For example

$$\mathcal{F}_5 = \left\{ \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{1}{1} \right\}$$

A Farey series is characterized by the property that if P_{i-1}/Q_{i-1} and P_i/Q_i are consecutive terms of $\mathcal{F}_Q$, then

$$P_i Q_{i-1} - P_{i-1} Q_i = 1.$$

The number of terms in the Farey series F_Q is given by

$$\Phi(Q) = \sum_{n=1}^Q \phi(n) = \frac{3Q^2}{\pi^2} + O(Q \ln Q). \quad (32.6)$$

Here the Euler function $\phi(Q)$ is the number of integers not exceeding and relatively prime to Q . For example, $\phi(1) = 1$, $\phi(2) = 1$, $\phi(3) = 2$, $\dots$, $\phi(12) = 4$, $\phi(13) = 12$, $\dots$

From a number-theorist’s point of view, the *continued fraction partitioning* of the unit interval is the most venerable organization of rationals, preferred already by Gauss. The continued fraction partitioning is obtained by ordering rationals corresponding to continued fractions of increasing length. If we turn this ordering into a way of covering the complementary set to mode-lockings in a circle map, then the first level is obtained by deleting $\Delta_{[1]}$, $\Delta_{[2]}$, $\dots$, $\Delta_{[a_1]}$, $\dots$ mode-lockings; their complement are the *covering* intervals $\ell_1, \ell_2, \dots, \ell_{a_1}, \dots$ which contain all windings, rational and irrational, whose continued fraction expansion starts with

$[a_1, \dots]$ and is of length at least 2. The second level is obtained by deleting $\Delta_{[1,2]}, \Delta_{[1,3]}, \dots, \Delta_{[2,2]}, \Delta_{[2,3]}, \dots, \Delta_{[n,m]}, \dots$ and so on.

The n th level continued fraction partition $\mathcal{S}_n = \{a_1 a_2 \dots a_n\}$ is defined as the monotonically increasing sequence of all rationals P_i/Q_i between 0 and 1 whose continued fraction expansion is of length n :

$$\frac{P_i}{Q_i} = [a_1, a_2, \dots, a_n] = \frac{1}{a_1 + \frac{1}{a_2 + \dots \frac{1}{a_n}}}$$

The object of interest, the set of the irrational winding numbers, is in this partitioning labeled by $\mathcal{S}_\infty = \{a_1 a_2 a_3 \dots\}$, $a_k \in \mathbb{Z}^+$, i.e., the set of winding numbers with infinite continued fraction expansions. The continued fraction labeling is particularly appealing in the present context because of the close connection of the Gauss shift to the renormalization transformation R , discussed below. The Gauss map

$$T(x) = \begin{cases} \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor & x \neq 0 \\ 0 & x = 0 \end{cases} \quad (32.7)$$

($\lfloor \cdot \rfloor$ denotes the integer part) acts as a shift on the continued fraction representation of numbers on the unit interval

$$x = [a_1, a_2, a_3, \dots] \rightarrow T(x) = [a_2, a_3, \dots] \quad (32.8)$$

into the “mother” interval $\ell_{a_2 a_3 \dots}$.

However natural the continued fractions partitioning might seem to a number theorist, it is problematic in practice, as it requires measuring infinity of mode-lockings even at the first step of the partitioning. Thus numerical and experimental use of continued fraction partitioning requires at least some understanding of the asymptotics of mode-lockings with large continued fraction entries.

The *Farey tree partitioning* is a systematic bisection of rationals: it is based on the observation that roughly halfway between any two large stability intervals (such as $1/2$ and $1/3$) in the devil’s staircase of figure 32.2 there is the next largest stability interval (such as $2/5$). The winding number of this interval is given by the Farey median $(P + P')/(Q + Q')$ of the parent mode-lockings P/Q and P'/Q' . This kind of cycle “gluing” is rather general and by no means restricted to circle maps; it can be attained whenever it is possible to arrange that the Q th iterate deviation caused by shifting a parameter from the correct value for the Q -cycle is exactly compensated by the Q' th iterate deviation from closing the Q' -cycle; in this way the two near cycles can be glued together into an exact cycle of period $Q + Q'$. The Farey tree is obtained by starting with the ends of the unit interval written as $0/1$ and $1/1$, and then recursively bisecting intervals by means of Farey mediants.

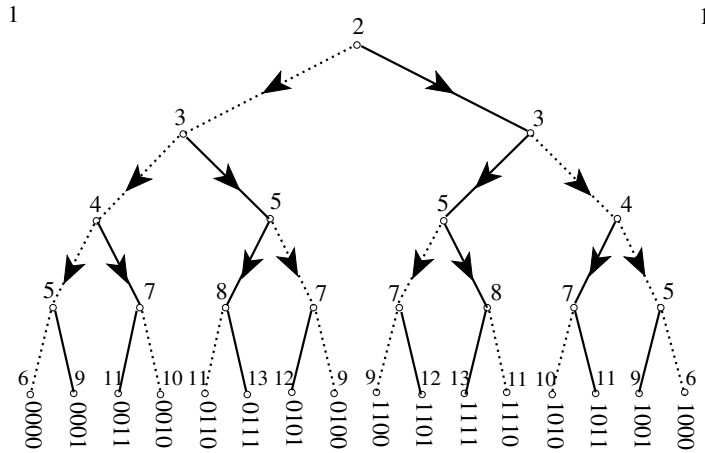


Figure 32.3: Farey tree: alternating binary ordered labeling of all Farey denominators on the n th Farey tree level.

We define the n th Farey tree level T_n as the monotonically increasing sequence of those continued fractions $[a_1, a_2, \dots, a_k]$ whose entries $a_i \geq 1$, $i = 1, 2, \dots, k-1$, $a_k \geq 2$, add up to $\sum_{i=1}^k a_i = n + 2$. For example

$$T_2 = \{[4], [2, 2], [1, 1, 2], [1, 3]\} = \left(\frac{1}{4}, \frac{1}{5}, \frac{3}{5}, \frac{3}{4}\right). \quad (32.9)$$

The number of terms in T_n is 2^n . Each rational in T_{n-1} has two “daughters” in T_n , given by

$$[\dots, a-1, 2] \quad [\dots, a] \quad [\dots, a+1]$$

Iteration of this rule places all rationals on a binary tree, labeling each by a unique binary label, figure 32.3.

The smallest and the largest denominator in T_n are respectively given by

$$[n-2] = \frac{1}{n-2}, \quad [1, 1, \dots, 1, 2] = \frac{F_{n+1}}{F_{n+2}} \propto \rho^n, \quad (32.10)$$

where the Fibonacci numbers F_n are defined by $F_{n+1} = F_n + F_{n-1}$; $F_0 = 0$, $F_1 = 1$, and ρ is the golden mean ratio

$$\rho = \frac{1 + \sqrt{5}}{2} = 1.61803\dots \quad (32.11)$$

Note the enormous spread in the cycle periods on the same level of the Farey tree: $n \leq Q \leq \rho^n$. The cycles whose period grows only as a power of the Farey tree level will cause strong nonhyperbolic effects in the evaluation of various averages.

Having defined the partitioning schemes of interest here, we now briefly summarize the results of the circle-map renormalization theory.

32.2 Local theory: “Golden mean” renormalization



The way to pinpoint a point on the border of order is to recursively adjust the parameters so that at the recurrence times $t = n_1, n_2, n_3, \dots$ the trajectory passes through a region of contraction sufficiently strong to compensate for the accumulated expansion of the preceding n_i steps, but not so strong as to force the trajectory into a stable attracting orbit. The *renormalization operation* R implements this procedure by recursively magnifying the neighborhood of a point on the border in the dynamical space (by rescaling by a factor α), in the parameter space (by shifting the parameter origin onto the border and rescaling by a factor δ), and by replacing the initial map f by the n th iterate f^n restricted to the magnified neighborhood

$$f_p(x) \rightarrow Rf_p(x) = \alpha f_{p/\delta}^n(x/\alpha)$$

There are by now many examples of such renormalizations in which the new function, framed in a smaller box, is a rescaling of the original function, i.e., the fix-point function of the renormalization operator R . The best known is the period doubling renormalization, with the recurrence times $n_i = 2^i$. The simplest circle map example is the golden mean renormalization, with recurrence times $n_i = F_i$ given by the Fibonacci numbers (32.10). Intuitively, in this context a metric self-similarity arises because iterates of critical maps are themselves critical, i.e., they also have cubic inflection points with vanishing derivatives.

The renormalization operator appropriate to circle maps acts as a generalization of the Gauss shift (32.38); it maps a circle map (represented as a pair of functions (g, f) , of winding number $[a, b, c, \dots]$ into a rescaled map of winding number $[b, c, \dots]$:

$$R_a \begin{pmatrix} g \\ f \end{pmatrix} = \begin{pmatrix} \alpha g^{a-1} \circ f \circ \alpha^{-1} \\ \alpha g^{a-1} \circ f \circ g \circ \alpha^{-1} \end{pmatrix}, \quad (32.12)$$

Acting on a map with winding number $[a, a, a, \dots]$, R_a returns a map with the same winding number $[a, a, \dots]$, so the fixed point of R_a has a quadratic irrational winding number $W = [a, a, a, \dots]$. This fixed point has a single expanding eigenvalue δ_a . Similarly, the renormalization transformation $R_{a_p} \dots R_{a_2} R_{a_1} \equiv R_{a_1 a_2 \dots a_p}$ has a fixed point of winding number $W_p = [a_1, a_2, \dots, a_{n_p}, a_1, a_2, \dots]$, with a single expanding eigenvalue δ_p .

For short repeating blocks, δ can be estimated numerically by comparing successive continued fraction approximants to W . Consider the P_r/Q_r rational approximation to a quadratic irrational winding number W_p whose continued fraction expansion consists of r repeats of a block p . Let Ω_r be the parameter for which the map (32.1) has a superstable cycle of rotation number $P_r/Q_r = [p, p, \dots, p]$. The δ_p can then be estimated by extrapolating from

$$\Omega_r - \Omega_{r+1} \propto \delta_p^{-r}. \quad (32.13)$$

What this means is that the “devil’s staircase” of figure 32.2 is self-similar under magnification by factor δ_p around any quadratic irrational W_p .

The fundamental result of the renormalization theory (and the reason why all this is so interesting) is that the ratios of successive P_r/Q_r mode-locked intervals converge to *universal* limits. The simplest example of (32.13) is the sequence of Fibonacci number continued fraction approximants to the golden mean winding number $W = [1, 1, 1, \dots] = (\sqrt{5} - 1)/2$.

When global problems are considered, it is useful to have at least an idea on external scaling laws for mode-lockings. This is achieved, in a first analysis, by fixing the cycle period Q and describing the range of possible asymptotics.

For a given cycle period Q , it is found that the *narrowest* interval shrinks with a power law

$$\Delta_{1/Q} \propto Q^{-3} \quad (32.14)$$

For fixed Q the *widest* interval is bounded by $P/Q = F_{n-1}/F_n$, the n th continued fraction approximant to the *golden mean*. The intuitive reason is that the golden mean winding sits as far as possible from any short cycle mode-locking.

The golden mean interval shrinks with a universal exponent

$$\Delta_{P/Q} \propto Q^{-2\mu_1} \quad (32.15)$$

where $P = F_{n-1}$, $Q = F_n$ and μ_1 is related to the universal Shenker number δ_1 (32.13) and the golden mean (32.11) by

$$\mu_1 = \frac{\ln |\delta_1|}{2 \ln \rho} = 1.08218 \dots \quad (32.16)$$

The closeness of μ_1 to 1 indicates that the golden mean approximant mode-lockings barely feel the fact that the map is critical (in the $k=0$ limit this exponent is $\mu = 1$).

To summarize: for critical maps the spectrum of exponents arising from the circle maps renormalization theory is bounded from above by the harmonic scaling, and from below by the geometric golden-mean scaling:

$$3/2 > \mu_{m/n} \geq 1.08218 \dots \quad (32.17)$$

32.3 Global theory: Thermodynamic averaging

Consider the following average over mode-locking intervals (32.5):

$$\Omega(\tau) = \sum_{Q=1}^{\infty} \sum_{(P|Q)=1} \Delta_{P/Q}^{-\tau} \quad (32.18)$$

The sum is over all irreducible rationals P/Q , $P < Q$, and $\Delta_{P/Q}$ is the width of the parameter interval for which the iterates of a critical circle map lock onto a cycle of period Q , with winding number P/Q .

The qualitative behavior of (32.18) is easy to pin down. For sufficiently negative τ , the sum is convergent; in particular, for $\tau = -1$, $\Omega(-1) = 1$, as for the critical circle maps the mode-lockings fill the entire Ω range [48]. However, as τ increases, the contributions of the narrow (large Q) mode-locked intervals $\Delta_{P/Q}$ get blown up to $1/\Delta_{P/Q}^\tau$, and at some critical value of τ the sum diverges. This occurs for $\tau < 0$, as $\Omega(0)$ equals the number of all rationals and is clearly divergent.

The sum (32.18) is infinite, but in practice the experimental or numerical mode-locked intervals are available only for small finite Q . Hence it is necessary to split up the sum into subsets $\mathcal{S}_n = \{i\}$ of rational winding numbers P_i/Q_i on the “level” n , and present the set of mode-lockings hierarchically, with resolution increasing with the level:

$$\bar{Z}_n(\tau) = \sum_{i \in \mathcal{S}_n} \Delta_i^{-\tau}. \quad (32.19)$$

The original sum (32.18) can now be recovered as the $z = 1$ value of a “generating” function $\Omega(z, \tau) = \sum_n z^n \bar{Z}_n(\tau)$. As z is anyway a formal parameter, and n is a rather arbitrary “level” in some *ad hoc* partitioning of rational numbers, we bravely introduce a still more general, P/Q weighted generating function for (32.18):

$$\Omega(q, \tau) = \sum_{Q=1}^{\infty} \sum_{(P|Q)=1} e^{-q\nu_{P/Q}} Q^{2\tau\mu_{P/Q}}. \quad (32.20)$$

The sum (32.18) corresponds to $q = 0$. Exponents $\nu_{P/Q}$ will reflect the importance we assign to the P/Q mode-locking, i.e., the *measure* used in the averaging over all mode-lockings. Three choices of the $\nu_{P/Q}$ hierarchy that we consider here correspond respectively to the Farey series partitioning

$$\Omega(q, \tau) = \sum_{Q=1}^{\infty} \Phi(Q)^{-q} \sum_{(P|Q)=1} Q^{2\tau\mu_{P/Q}}, \quad (32.21)$$

the continued fraction partitioning

$$\Omega(q, \tau) = \sum_{n=1}^{\infty} e^{-qn} \sum_{[a_1, \dots, a_n]} Q^{2\tau\mu_{[a_1, \dots, a_n]}}, \quad (32.22)$$

and the Farey tree partitioning

$$\Omega(q, \tau) = \sum_{k=n}^{\infty} 2^{-qn} \sum_{i=1}^{2^n} Q_i^{2\tau\mu_i}, \quad Q_i/P_i \in T_n. \quad (32.23)$$

We remark that we are investigating a set arising in the analysis of the parameter space of a dynamical system: there is no “natural measure” dictated by dynamics, and the choice of weights reflects only the choice of hierarchical presentation.

32.4 Hausdorff dimension of irrational windings

A finite cover of the set irrational windings at the “ n th level of resolution” is obtained by deleting the parameter values corresponding to the mode-lockings in the subset $\mathcal{S}_n$; left behind is the set of complement *covering* intervals of widths

$$\ell_i = \Omega_{P_r/Q_r}^{\min} - \Omega_{P_l/Q_l}^{\max} . \quad (32.24)$$

Here $\Omega_{P_r/Q_r}^{\min}$ ($\Omega_{P_l/Q_l}^{\max}$) are respectively the lower (upper) edges of the mode-locking intervals Δ_{P_r/Q_r} (Δ_{P_l/Q_l}) bounding ℓ_i and i is a symbolic dynamics label, for example the entries of the continued fraction representation $P/Q = [a_1, a_2, \dots, a_n]$ of one of the boundary mode-lockings, $i = a_1 a_2 \dots a_n$. ℓ_i provide a finite cover for the irrational winding set, so one may consider the sum

$$Z_n(\tau) = \sum_{i \in \mathcal{S}_n} \ell_i^{-\tau} \quad (32.25)$$

The value of $-\tau$ for which the $n \rightarrow \infty$ limit of the sum (32.25) is finite is the *Hausdorff dimension* D_H of the irrational winding set. Strictly speaking, this is the Hausdorff dimension only if the choice of covering intervals ℓ_i is optimal; otherwise it provides an upper bound to D_H . As by construction the ℓ_i intervals cover the set of irrational winding with no slack, we expect that this limit yields the Hausdorff dimension. This is supported by all numerical evidence, but a proof that would satisfy mathematicians is lacking.

The physically relevant statement is that for critical circle maps $D_H = 0.870 \dots$ is a (global) universal number.

exercise 32.2

32.4.1 The Hausdorff dimension in terms of cycles

Estimating the $n \rightarrow \infty$ limit of (32.25) from finite numbers of covering intervals ℓ_i is a rather unilluminating chore. Fortunately, there exist considerably more elegant ways of extracting D_H . We have noted that in the case of the “trivial” mode-locking problem (32.4), the covering intervals are generated by iterations of the Farey map (32.37) or the Gauss shift (32.38). The n th level sum (32.25) can be approximated by $\mathcal{L}_\tau^n$, where

$$\mathcal{L}_\tau(y, x) = \delta(x - f^{-1}(y)) |f'(y)|^\tau$$

This amounts to approximating each cover width ℓ_i by $|df^n/dx|$ evaluated on the i th interval. We are thus led to the following determinant

$$\begin{aligned} \det(1 - \mathcal{L}_\tau) &= \exp \left(- \sum_p \sum_{r=1}^{\infty} \frac{z^{rn_p}}{r} \frac{|\Lambda_p^r|^\tau}{1 - 1/\Lambda_p^r} \right) \\ &= \prod_p \prod_{k=0}^{\infty} \left(1 - z^{n_p} |\Lambda_p|^\tau / \Lambda_p^k \right) . \end{aligned} \quad (32.26)$$

The sum (32.25) is dominated by the leading eigenvalue of $\mathcal{L}_\tau$; the Hausdorff dimension condition $Z_n(-D_H) = O(1)$ means that $\tau = -D_H$ should be such that the leading eigenvalue is $z = 1$. The leading eigenvalue is determined by the $k = 0$ part of (32.26); putting all these pieces together, we obtain a pretty formula relating the Hausdorff dimension to the prime cycles of the map $f(x)$:

$$0 = \prod_p \left(1 - 1/|\Lambda_p|^{D_H}\right). \quad (32.27)$$

For the Gauss shift (32.38) the stabilities of periodic cycles are available analytically, as roots of quadratic equations: For example, the x_a fixed points (quadratic irrationals with $x_a = [a, a, a, \dots]$ infinitely repeating continued fraction expansion) are given by

$$x_a = \frac{-a + \sqrt{a^2 + 4}}{2}, \quad \Lambda_a = -\left(\frac{a + \sqrt{a^2 + 4}}{2}\right)^2 \quad (32.28)$$

and the $x_{ab} = [a, b, a, b, a, b, \dots]$ 2-cycles are given by

$$\begin{aligned} x_{ab} &= \frac{-ab + \sqrt{(ab)^2 + 4ab}}{2b} \\ \Lambda_{ab} &= (x_{ab}x_{ba})^{-2} = \left(\frac{ab + 2 + \sqrt{ab(ab + 4)}}{2}\right)^2 \end{aligned} \quad (32.29)$$

We happen to know beforehand that $D_H = 1$ (the irrationals take the full measure on the unit interval, or, from another point of view, the Gauss map is not a repeller), so is the infinite product (32.27) merely a very convoluted way to compute the number 1? Possibly so, but once the meaning of (32.27) has been grasped, the corresponding formula for the *critical* circle maps follows immediately:

$$0 = \prod_p \left(1 - 1/|\delta_p|^{D_H}\right). \quad (32.30)$$

The importance of this formula relies on the fact that it expresses D_H in terms of *universal* quantities, thus providing a nice connection from local universal exponents to global scaling quantities: actual computations using (32.30) are rather involved, as they require a heavy computational effort to extract Shenker's scaling δ_p for periodic continued fractions, and moreover dealing with an infinite alphabet requires control over tail summation if an accurate estimate is to be sought. In table 32.1 we give a small selection of computed Shenker's scalings.

32.5 Thermodynamics of Farey tree: Farey model



We end this chapter by giving an example of a number theoretical model motivated by the mode-locking phenomenology. We will consider it by means of the thermodynamic formalism of appendix A32, by looking at the free energy.

Table 32.1: Shenker's δ_p for a few periodic continued fractions, from ref. [9].

p	δ_p
[1 1 1 1 ...]	-2.833612
[2 2 2 2 ...]	-6.7992410
[3 3 3 3 ...]	-13.760499
[4 4 4 4 ...]	-24.62160
[5 5 5 5 ...]	-40.38625
[6 6 6 6 ...]	-62.140
[1 2 1 2 ...]	17.66549
[1 3 1 3 ...]	31.62973
[1 4 1 4 ...]	50.80988
[1 5 1 5 ...]	76.01299
[2 3 2 3 ...]	91.29055

Consider the Farey tree partition sum (32.23): the narrowest mode-locked interval (32.15) at the n th level of the Farey tree partition sum (32.23) is the golden mean interval

$$\Delta_{F_{n-1}/F_n} \propto |\delta_1|^{-n}. \quad (32.31)$$

It shrinks exponentially, and for τ positive and large it dominates $q(\tau)$ and bounds $dq(\tau)/d\tau$:

$$q'_{max} = \frac{\ln |\delta_1|}{\ln 2} = 1.502642 \dots \quad (32.32)$$

However, for τ large and negative, $q(\tau)$ is dominated by the interval (32.14) which shrinks only harmonically, and $q(\tau)$ approaches 0 as

$$\frac{q(\tau)}{\tau} = \frac{3 \ln n}{n \ln 2} \rightarrow 0. \quad (32.33)$$

So for finite n , $q_n(\tau)$ crosses the τ axis at $-\tau = D_n$, but in the $n \rightarrow \infty$ limit, the $q(\tau)$ function exhibits a phase transition; $q(\tau) = 0$ for $\tau < -D_H$, but is a non-trivial function of τ for $-D_H \leq \tau$. This non-analyticity is rather severe - to get a clearer picture, we illustrate it by a few number-theoretic models (the critical circle maps case is qualitatively the same).

An approximation to the “trivial” Farey level thermodynamics is given by the “Farey model,” in which the intervals $\ell_{P/Q}$ are replaced by Q^{-2} :

$$Z_n(\tau) = \sum_{i=1}^{2^n} Q_i^{2\tau}. \quad (32.34)$$

Here Q_i is the denominator of the i th Farey rational P_i/Q_i . For example (see figure 32.3),

$$Z_2(1/2) = 4 + 5 + 5 + 4.$$

By the annihilation property (32.38) of the Gauss shift on rationals, the n th Farey level sum $Z_n(-1)$ can be written as the integral

$$Z_n(-1) = \int dx \delta(f^n(x)) = \sum 1/|f'_{a_1 \dots a_k}(0)|,$$

$\tau/2$	$Z_n(\tau/2)/Z_{n-1}(\tau/2)$
0	2
1	3
2	$(5 + \sqrt{17})/2$
3	7
4	$(5 + \sqrt{17})/2$
5	$7 + 4\sqrt{6}$
6	26.20249...

Table 32.2: Partition function sum rules for the Farey model.

and in general

$$Z_n(\tau) = \int dx \mathcal{L}_\tau^n(0, x),$$

with the sum restricted to the Farey level $a_1 + \dots + a_k = n + 2$. It is easily checked that $f'_{a_1 \dots a_k}(0) = (-1)^k Q^2_{[a_1, \dots, a_k]}$, so the Farey model sum is a partition generated by the Gauss map preimages of $x = 0$, i.e., by rationals, rather than by the quadratic irrationals as in (32.26). The sums are generated by the same transfer operator, so the eigenvalue spectrum should be the same as for the periodic orbit expansion, but in this variant of the finite level sums we can evaluate $q(\tau)$ *exactly* for $\tau = k/2$, k a nonnegative integer. First, one observes that $Z_n(0) = 2^n$. It is also easy to check that $Z_n(1/2) = \sum_i Q_i = 2 \cdot 3^n$. More surprisingly, $Z_n(3/2) = \sum_i Q_i^3 = 54 \cdot 7^{n-1}$. A few of these “sum rules” are listed in the table 32.2, they are consequence of the fact that the denominators on a given level are Farey sums of denominators on preceding levels.

exercise 32.3

A bound on D_H can be obtained by approximating (32.34) by

$$Z_n(\tau) = n^{2\tau} + 2^n \rho^{2n\tau}. \quad (32.35)$$

In this approximation we have replaced all $\ell_{P/Q}$, except the widest interval $\ell_{1/n}$, by the narrowest interval ℓ_{F_{n-1}/F_n} (see (32.15)). The crossover from the harmonic dominated to the golden mean dominated behavior occurs at the τ value for which the two terms in (32.35) contribute equally:

$$D_n = \hat{D} + O\left(\frac{\ln n}{n}\right), \quad \hat{D} = \frac{\ln 2}{2 \ln \rho} = .72 \dots \quad (32.36)$$

For negative τ the sum (32.35) is the lower bound on the sum (32.25), so $\hat{D}$ is a lower bound on D_H .

From a general perspective the analysis of circle maps thermodynamics has revealed the fact that physically interesting dynamical systems often exhibit mixtures of hyperbolic and marginal stabilities. In such systems there are orbits that stay ‘glued’ arbitrarily close to stable regions for arbitrarily long times. This is a generic phenomenon for Hamiltonian systems, where elliptic islands of stability coexist with hyperbolic homoclinic tangles. Thus the considerations of chapter 29 are important also in the analysis of renormalization at the onset of chaos.

Résumé

The mode locking problem, and the quasiperiodic transition to chaos offer an opportunity to use cycle expansions on hierarchical structures in parameter space: this is not just an application of the conventional thermodynamic formalism, but offers a clue on how to extend universality theory from local scalings to global quantities.

Commentary

Remark 32.1. The physics of circle maps. Mode-locking phenomenology is reviewed in ref. [18], a more theoretically oriented discussion is contained in ref. [21]. While representative of dissipative systems we may also consider circle maps as a crude approximation to Hamiltonian local dynamics: a typical island of stability in a Hamiltonian 2-dimensional map is an infinite sequence of concentric KAM tori and chaotic regions. In the crudest approximation, the radius can here be treated as an external parameter Ω , and the angular motion can be modeled by a map periodic in the angular variable [45, 46]. By losing all of the ‘island-within-island’ structure of real systems, circle map models skirt the problems of determining the symbolic dynamics for a realistic Hamiltonian system, but they do retain some of the essential features of such systems, such as the golden mean renormalization [19, 46] and nonhyperbolicity in form of sequences of cycles accumulating toward the borders of stability. In particular, in such systems there are orbits that stay “glued” arbitrarily close to stable regions for arbitrarily long times. As this is a generic phenomenon in physically interesting dynamical systems, such as the Hamiltonian systems with coexisting elliptic islands of stability and hyperbolic homoclinic tangles, development of good computational techniques is here of utmost practical importance.

Remark 32.2. Critical mode-locking set The fact that mode-lockings completely fill the unit interval at the critical point has been proposed in refs. [21, 30]. The proof that the set of irrational windings is of zero Lebesgue measure is given in ref. [48].

Remark 32.3. Counting noise for Farey series. The number of rationals in the Farey series of order Q is $\phi(Q)$, which is a highly irregular function of Q : incrementing Q by 1 increases $\Phi(Q)$ by anything from 2 to Q terms. We refer to this fact as the “Euler noise.”

The Euler noise poses a serious obstacle for numerical calculations with the Farey series partitionings; it blocks smooth extrapolations to $Q \rightarrow \infty$ limits from finite Q data. While this in practice renders inaccurate most Farey-sequence partitioned averages, the finite Q Hausdorff dimension estimates exhibit (for reasons that we do not understand) surprising numerical stability, and the Farey series partitioning actually yields the *best* numerical value of the Hausdorff dimension (32.25) of any methods used so far; for example the computation in ref. [3] for critical sine map (32.1), based on $240 \leq Q \leq 250$ Farey series partitions, yields $D_H = .87012 \pm .00001$. The quoted error refers to the variation of D_H over this range of Q ; as the computation is not asymptotic, such numerical stability can underestimate the actual error by a large factor.

Remark 32.4. Farey tree presentation function. The Farey tree rationals can be generated by backward iterates of $1/2$ by the Farey presentation function [15]:

$$\begin{aligned} f_0(x) &= x/(1-x) & 0 \leq x < 1/2 \\ f_1(x) &= (1-x)/x & 1/2 < x \leq 1. \end{aligned} \quad (32.37)$$

The Gauss shift (32.7) corresponds to replacing the binary Farey presentation function branch f_0 in (32.37) by an infinity of branches

$$\begin{aligned} f_a(x) &= f_1 \circ f_0^{(a-1)}(x) = \frac{1}{x} - a, & \frac{1}{a-1} < x \leq \frac{1}{a}, \\ f_{ab\dots c}(x) &= f_c \circ \dots \circ f_b \circ f_a(x). \end{aligned} \quad (32.38)$$

A rational $x = [a_1, a_2, \dots, a_k]$ is annihilated by the k th iterate of the Gauss shift, $f_{a_1 a_2 \dots a_k}(x) = 0$. The above maps look innocent enough, but note that what is being partitioned is not the dynamical space, but the parameter space. The flow described by (32.37) and by its non-trivial circle-map generalizations will turn out to be a *renormalization group* flow in the function space of dynamical systems, not an ordinary flow in the state space of a particular dynamical system.

The Farey tree has a variety of interesting symmetries (such as “flipping heads and tails” relations obtained by reversing the order of the continued-fraction entries) with as yet unexploited implications for the renormalization theory: some of these are discussed in ref. [11].

An alternative labeling of Farey denominators has been introduced by Knauf [29] in context of number-theoretical modeling of ferromagnetic spin chains: it allows for a number of elegant manipulations in thermodynamic averages connected to the Farey tree hierarchy.

Remark 32.5. Circle map renormalization The idea underlying golden mean renormalization goes back to Shenker [45]. A renormalization group procedure was formulated in refs. [16, 40, 44], where moreover the uniqueness of the relevant eigenvalue is claimed. This statement has been confirmed by a computer-assisted proof [37], and in the following we will always assume it. There are a number of experimental evidences for local universality, see refs. [20, 47].

On the other side of the scaling tale, the power law scaling for harmonic fractions (discussed in refs. [11, 22–24]) is derived by methods akin to those used in describing intermittency [42]: $1/Q$ cycles accumulate toward the edge of $0/1$ mode-locked interval, and as the successive mode-locked intervals $1/Q$, $1/(Q-1)$ lie on a parabola, their differences are of order Q^{-3} .

Remark 32.6. Farey series and the Riemann hypothesis The Farey series thermodynamics is of a number theoretical interest, because the Farey series provide uniform coverings of the unit interval with rationals, and because they are closely related to the deepest problems in number theory, such as the Riemann hypothesis [13, 49]. The distribution of the Farey series rationals across the unit interval is surprisingly uniform - indeed, so uniform that in the pre-computer days it has motivated a compilation of an entire handbook of Farey series [38]. A quantitative measure of the non-uniformity of the distribution of Farey rationals is given by displacements of Farey rationals for $P_i/Q_i \in \mathcal{F}_Q$ from uniform spacing:

$$\delta_i = \frac{i}{\Phi(Q)} - \frac{P_i}{Q_i}, \quad i = 1, 2, \dots, \Phi(Q)$$

The Riemann hypothesis states that the zeros of the Riemann zeta function lie on the $s = 1/2 + i\tau$ line in the complex s plane, and would seem to have nothing to do with physicists' real mode-locking widths that we are interested in here. However, there is a real-line version of the Riemann hypothesis that lies very close to the mode-locking problem. According to the theorem of Franel and Landau [13, 17, 49], the Riemann hypothesis is equivalent to the statement that

$$\sum_{Q_i \leq Q} |\delta_i| = o(Q^{\frac{1}{2} + \epsilon})$$

for all ϵ as $Q \rightarrow \infty$. The mode-lockings $\Delta_{P/Q}$ contain the necessary information for constructing the partition of the unit interval into the ℓ_i covers, and therefore implicitly contain the δ_i information. The implications of this for the circle-map scaling theory have not been worked out, and is not known whether some conjecture about the thermodynamics of irrational windings is equivalent to (or harder than) the Riemann hypothesis, but the danger lurks.

Remark 32.7. Farey tree partitioning. The Farey tree partitioning was introduced in refs. [10, 11, 51] and its thermodynamics is discussed in detail in refs. [3, 15]. The Farey tree hierarchy of rationals is rather new, and, as far as we are aware, not previously studied by number theorists. It is appealing both from the experimental and from the golden-mean renormalization point of view, but it has a serious drawback of lumping together mode-locking intervals of wildly different sizes on the same level of the Farey tree.

Remark 32.8. Local and global universality. Numerical evidences for global universal behavior have been presented in ref. [21]. The question was reexamined in ref. [3], where it was pointed out how a high-precision numerical estimate is in practice very hard to obtain. It is not at all clear whether this is the optimal global quantity to test but at least the Hausdorff dimension has the virtue of being independent of how one partitions mode-lockings and should thus be the same for the variety of thermodynamic averages in the literature.

The formula (32.30), linking local to global behavior, was proposed in ref. [9].

The derivation of (32.30) relies only on the following aspects of the "hyperbolicity conjecture" of refs. [11, 14, 27, 31, 43]:

1. *limits* for Shenker δ 's *exist* and are universal. This should follow from the renormalization theory developed in refs. [16, 37, 40, 44], though a general proof is still lacking.
2. δ_p grow *exponentially* with n_p , the length of the continued fraction block p .
3. δ_p for $p = a_1 a_2 \dots n$ with a large continued fraction entry n grows as a *power* of n . According to (32.14), $\lim_{n \rightarrow \infty} \delta_p \propto n^3$. In the calculation of ref. [9] the explicit values of the asymptotic exponents and prefactors were not used, only the assumption that the growth of δ_p with n is not slower than a power of n .

Remark 32.9. Farey model. The Farey model (32.33) has been proposed in ref. [3]; though it might seem to have been pulled out of a hat, the Farey model is as sensible description of the distribution of rationals as the periodic orbit expansion (32.26).

Remark 32.10. Symbolic dynamics for Hamiltonian rotational orbits. The rotational codes of ref. [12] are closely related to those for maps with a natural angle variable, for

example for circle maps [50, 52] and cat maps [41]. Ref. [12] also offers a systematic rule for obtaining the symbolic codes of “islands around islands” rotational orbits [35]. These correspond, for example, to orbits that rotate around orbits that rotate around the elliptic fixed point; thus they are defined by a sequence of rotation numbers.

A different method for constructing symbolic codes for “islands around islands” was given in refs. [1, 2]; however in these cases the entire set of orbits in an island was assigned the same sequence and the motivation was to study the transport implications for chaotic orbits outside the islands [35, 36].

Remark 32.11. Three-frequency mode locking. P. Cvitanović (notes available upon request) has extended the two-frequency mode-locking golden-mean renormalization to the three-frequency mode-locking ‘spiral mean’ renormalization theory, where the golden and metal means (solutions of quadratic equations) are generalized to Pisot–Vijayaraghavan (PV) numbers (solutions of cubic or higher order equations). Kim and Ostlund discuss this problem in refs. [25, 26, 28]. See also refs. [4–6, 33].

References

- [1] V. Afraimovich, A. Maass, and J. Urías, “Symbolic dynamics for sticky sets in Hamiltonian systems”, *Nonlinearity* **13**, 617–637 (2000).
- [2] Y. Aizawa, “Symbolic dynamics approach to the two-D chaos in area-preserving maps”, *Progr. Theor. Phys.* **71**, 1419–1421 (1984).
- [3] R. Artuso, P. Cvitanović, and B. G. Kenny, “Phase transitions on strange irrational sets”, *Phys. Rev. A* **39**, 268–281 (1989).
- [4] C. Baesens, J. Guckenheimer, S. Kim, and R. S. MacKay, “Three coupled oscillators: mode-locking, global bifurcations and toroidal chaos”, *Physica D* **49**, 387–475 (1991).
- [5] P. R. Baldwin, “A multidimensional continued fraction and some of its statistical properties”, *J. Stat. Phys.* **66**, 1463–1505 (1992).
- [6] S.-g. Chen and Y.-q. Wang, “Cubic irrational number and critical scaling law for quasiperiodic motion”, *Phys. Lett. A* **153**, 113–116 (1991).
- [7] P. Contucci and A. Knauf, “The phase transition of the number-theoretical spin chain”, *Forum Mathematicum* **9**, 547–568 (1997).
- [8] A. Csordás and P. Szépfalusy, “Dynamical multifractal properties of a map related to a chaotic cosmological model”, *Phys. Rev. A* **40**, 2221–2224 (1989).
- [9] P. Cvitanović, G. H. Gunaratne, and M. J. Vinson, “On the mode-locking universality for critical circle maps”, *Nonlinearity* **3**, 873–885 (1990).
- [10] P. Cvitanović and J. Myrheim, “Universality for period n -tuplings in complex mappings”, *Phys. Lett. A* **94**, 329–333 (1983).
- [11] P. Cvitanović, B. Shraiman, and B. Söderberg, “Scaling laws for mode lockings in circle maps”, *Phys. Scr.* **32**, 263–270 (1985).

- [12] H. R. Dullin, J. D. Meiss, and D. G. Sterling, “Symbolic codes for rotational orbits”, *SIAM J. Appl. Dyn. Sys.* **4**, 515–562 (2005).
- [13] H. M. Edwards, *Riemann’s Zeta Function* (Dover, New York, 2003).
- [14] M. J. Feigenbaum, “Presentation functions and scaling function theory for circle maps”, *Nonlinearity* **1**, 577–602 (1988).
- [15] M. J. Feigenbaum, “Presentation functions, fixed points, and a theory of scaling function dynamics”, *J. Stat. Phys.* **52**, 527–569 (1988).
- [16] M. J. Feigenbaum, L. P. Kadanoff, and S. J. Shenker, “Quasiperiodicity in dissipative systems: A renormalization group analysis”, *Physica D* **5**, 370–386 (1982).
- [17] J. Franel, “Les suites de Farey et le problème des nombres premiers”, *Nachr. Ges. Wiss. Gottingen, Math.-Phys.* **1924**, 198–201 (1924).
- [18] J. Glazier and A. Libchaber, “Quasi-periodicity and dynamical systems: An experimentalist’s view”, *IEEE Trans. Circuits and Systems* **35**, 790–809 (1988).
- [19] J. M. Greene, “A method for determining a stochastic transition”, *J. Math. Phys.* **20**, 1183–1201 (1979).
- [20] E. G. Gwinn and R. M. Westervelt, “Scaling structure of attractors at the transition from quasiperiodicity to chaos in electronic transport in Ge”, *Phys. Rev. Lett.* **59**, 157–160 (1987).
- [21] M. H. Jensen, P. Bak, and T. Bohr, “Complete devil’s staircase fractal dimension and universality of mode-locking structure in the circle map”, *Phys. Rev. Lett.* **50**, 1637–1639 (1983).
- [22] K. Kaneko, “On the period-adding phenomena at the frequency locking in a one-dimensional mapping”, *Progr. Theor. Phys.* **68**, 669–672 (1982).
- [23] K. Kaneko, “Similarity structure and scaling property of the period-adding phenomena”, *Progr. Theor. Phys.* **69**, 403–414 (1983).
- [24] K. Kaneko, “Transition from torus to chaos accompanied by frequency lockings with symmetry breaking: In connection with the coupled-logistic map”, *Progr. Theor. Phys.* **69**, 1427–1442 (1983).
- [25] S. Kim and S. Ostlund, “Renormalization of mappings of the two-torus”, *Phys. Rev. Lett.* **55**, 1165–1168 (1985).
- [26] S. Kim and S. Ostlund, “Simultaneous rational approximations in the study of dynamical systems”, *Phys. Rev. A* **34**, 3426–3434 (1986).
- [27] S. Kim and S. Ostlund, “Universal scaling in circle maps”, *Physica D* **39**, 365–392 (1989).
- [28] S. Kim, S. Ostlund, and G. Yu, “Fourier analysis of multi-frequency dynamical systems”, *Physica D* **31**, 117–126 (1988).
- [29] A. Knauf, “On a ferromagnetic spin chain”, *Commun. Math. Phys.* **153**, 77–115 (1993).
- [30] O. E. Lanford, “A numerical study of the likelihood of phase locking”, *Physica D* **14**, 403–408 (1985).

- [31] O. E. Lanford, “Renormalisation analysis of critical circle mapping with general rotation number”, in *VIII International Congress on Mathematical Physics*, edited by M. Mebkhout and R. S  n  or (World Scientific, Singapore, 1987), pp. 532–536.
- [32] D. Levin, “Development of non-linear transformations for improving convergence of sequences”, *Int. J. Comput. Math.* **3**, 371–388 (1972).
- [33] Z. Ji-Lin, B. Hu, and S. Yi-Sui, “Universal behaviour on the break-up of the spiral mean torus”, *Chin. Phys. Lett.* **18**, 1550–1553 (2001).
- [34] D. Mayer and G. Roepstorff, “On the relaxation time of Gauss’s continued-fraction map I. The Hilbert space approach (Koopmanism)”, *J. Stat. Phys.* **47**, 149–171 (1987).
- [35] J. D. Meiss, “Class renormalization: Islands around islands”, *Phys. Rev. A* **34**, 2375–2383 (1986).
- [36] J. D. Meiss and E. Ott, “Markov tree model of transport in area-preserving maps”, *Physica D* **20**, 387–402 (1986).
- [37] B. D. Mestel, Computer assisted proof of universality of cubic critical maps of the circle with golden mean rotation number, PhD thesis (University of Warwick, 1985).
- [38] E. H. Neville, *The Farey Series of Order 1025* (Cambridge Univ. Press, Cambridge UK, 1950).
- [39] N. Osada, “A convergence acceleration method for some logarithmically convergent sequences”, *SIAM J. Numer. Anal.* **27**, 178–189 (1990).
- [40] S. Ostlund, D. Rand, J. Sethna, and E. Siggia, “Universal properties of the transition from quasi-periodicity to chaos in dissipative systems”, *Physica D* **8**, 303–342 (1983).
- [41] I. Percival and F. Vivaldi, “Arithmetical properties of strongly chaotic motions”, *Physica D* **25**, 105–130 (1987).
- [42] Y. Pomeau and P. Manneville, “Intermittent transition to turbulence in dissipative dynamical systems”, *Commun. Math. Phys.* **74**, 189 (1980).
- [43] D. A. Rand, “Global phase space universality, smooth conjugacies and renormalisation. I. The $C^{1+\alpha}$ case”, *Nonlinearity* **1**, 181–202 (1988).
- [44] D. Rand, S. Ostlund, J. Sethna, and E. D. Siggia, “Universal transition from quasiperiodicity to chaos in dissipative systems”, *Phys. Rev. Lett.* **49**, 132–135 (1982).
- [45] S. J. Shenker, “Scaling behavior in a map of a circle onto itself: Empirical results”, *Physica D* **5**, 405–411 (1982).
- [46] S. J. Shenker and L. P. Kadanoff, “Critical behavior of a KAM surface: I. Empirical results”, *J. Stat. Phys.* **27**, 631–656 (1982).
- [47] J. Stavans, F. Heslot, and A. Libchaber, “Fixed winding number and the quasiperiodic route to chaos in a convective fluid”, *Phys. Rev. Lett.* **55**, 596–599 (1985).
- [48] G.   wi  tek, “Rational rotation numbers for maps of the circle”, *Commun. Math. Phys.* **119**, 109–128 (1988).

- [49] E. C. Titchmarsh and D. R. Heath-Brown, *The Theory of the Riemann Zeta-function* (Clarendon Press, Oxford UK, 1986).
- [50] P. Veerman, “Symbolic dynamics and rotation numbers”, *Physica A* **134**, 543–576 (1986).
- [51] G. T. Williams and D. H. Browne, “A family of integers and a theorem on circles”, *Amer. Math. Monthly* **54**, 534 (1947).
- [52] W.-M. Zheng, “Symbolic dynamics for the circle map”, *Int. J. Mod. Phys. B* **5**, 481–495 (1991).

32.6 Exercises

- 32.1. **Mode-locked intervals.** Check that when $k \neq 0$ the interval $\Delta_{P/Q}$ have a non-zero width (look for instance at simple fractions, and consider k small). Show that for small k the width of $\Delta_{0/1}$ is an increasing function of k .
- 32.2. **Bounds on Hausdorff dimension.** By making use of the bounds (32.17) show that the Hausdorff dimension for critical mode lockings may be bounded by

$$2/3 \leq D_H \leq .9240 \dots$$

- 32.3. **Farey model sum rules.** Verify the sum rules reported in table 32.2. An elegant way to get a number of sum rules for the Farey model is by taking into account an lexical ordering introduced by Contucci and Knauf [7].
- 32.4. **Metric entropy of the Gauss shift.** Check that the Lyapunov exponent of the Gauss map (32.7) is given by $\pi^2/6 \ln 2$. This result has been claimed to be relevant in the discussion of "mixmaster" cosmologies, see ref. [8].
- 32.5. **Refined expansions.** Show that the above estimates can be refined as follows:

$$F(z, 2) \sim \zeta(2) + (1 - z) \log(1 - z) - (1 - z)$$

and

$$F(z, s) \sim \zeta(s) + \Gamma(1 - s)(1 - z)^{s-1} - S(s)(1 - z)$$

for $s \in (1, 2)$ ($S(s)$ being expressed by a converging sum). You may use either more detailed estimate for $\zeta(s, a)$ (via Euler summation formula) or keep on subtracting leading contributions.

- 32.6. j_n and α_{cr} . Look at the integration region and how it scales by plotting it for increasing values of n .
- 32.7. **Estimates of the Riemann zeta function.** Try to approximate numerically the Riemann zeta function for $s = 2, 4, 6$ using different acceleration algorithms: check your results with refs. [32, 39].
- 32.8. **Farey tree and continued fractions I.** Consider the Farey tree presentation function $f : [0, 1] \mapsto [0, 1]$, such that if $I = [0, 1/2)$ and $J = [1/2, 1]$, $f|_I = x/(1 - x)$ and $f|_J = (1 - x)/x$. Show that the corresponding induced map is the Gauss map $g(x) = 1/x - [1/x]$.
- 32.9. **Farey tree and continued fraction II. (Lethal weapon II).** Build the simplest piecewise linear approximation to the Farey tree presentation function (hint: substitute first the rightmost, hyperbolic branch with a linear one): consider then the spectral determinant of the induced map $\hat{g}$, and calculate the first two eigenvalues besides the probability conservation one. Compare the results with the rigorous bound deduced in ref. [34].

Part IV

The rest is noise

Q UNADRY: all these cycles, but how many do I need? Any physical system suffers background noise, any numerical prediction suffers computational roundoff noise, and any set of equations models nature up to a given accuracy, since degrees of freedom are always neglected. If the noise is weak, the short-time dynamics is not altered significantly: short periodic orbits of the deterministic flow still partition coarsely the state space.

1. What is “noise”? (chapter 33)
2. Variational principles of classical mechanics, and path integrals of quantum mechanics, reimaged (chapter 34)

Chapter 33

Noise

He who establishes his argument by noise and command shows that his reason is weak.

—M. de Montaigne

THIS CHAPTER (which reader can safely skip on the first reading) is about noise, how it affects deterministic dynamics, and the ways it mimics quantum dynamics.

Why - in a monograph on deterministic and quantum chaos - start discussing noise? First, in physical settings any dynamics takes place against a noisy background, and whatever prediction we might have, we have to check its robustness to noise. Second, as we show in this chapter, to the leading order in noise strength, the semiclassical Hamilton-Jacobi formalism applies to weakly stochastic flows in toto. As deterministic noisy dynamics is more intuitive than quantum dynamics, understanding effects of noise helps demystify some of the formal machinery of semiclassical quantization. Surprisingly, symplectic structure emerges here not as a deep principle of mechanics, but an artifact of the leading approximation to quantum/noisy dynamics. Third, the variational principle derived here turns out to be a powerful tool for determining periodic orbits, see chapter 34. And, last but not least, upon some reflection, the whole enterprise of replacing deterministic trajectories by deterministic evolution operators, chapters 19 to 23, seems fatally flawed; if we have given up infinite precision in specifying initial conditions, why do we allow ourselves the infinite precision in the specification of evolution laws, i.e., define the evolution operator by means of the Dirac delta function $\delta(y - f^t(x))$? It will be comforting to learn that the deterministic evolution operators survive unscathed, as the leading approximation to the noisy ones in the limit of weak noise.



section 38.1

Another key result derived here is the evolution law (33.45) for the covariance matrix Q_a of a linearly evolved Gaussian density,

$$Q_{a+1} = M_a Q_a M_a^\top + \Delta_a.$$

To keep things simple we shall describe covariance evolution in the discrete time dynamics context, but the results apply both to the continuous and discrete time

flows. The most important lesson, however, is that physicist's Brownian diffusion intuition -that the effect of the noise is to spread out the trajectory as $\sqrt{t}$ - is *wrong*: In nonlinear dynamics the noise is always *local*, determined by balancing local nonlinear dynamics against the memory of the noise past.

section 33.5

We start by deriving the continuity equation for purely deterministic, noiseless flow, and then incorporate noise in stages: diffusion equation, Langevin equation, Fokker-Planck equation, stochastic path integrals, Hamilton-Jacobi formulation.

33.1 Deterministic transport

(E.A. Spiegel and P. Cvitanović)

The large body of accrued wisdom on the subject of flows called fluid dynamics is about physical flows of media with continuous densities. On the other hand, the flows in state spaces of dynamical systems frequently require more abstract tools. To sharpen our intuition about those, it is helpful to outline the more tangible fluid dynamical vision.

Consider first the simplest property of a fluid flow called *material invariant*. A material invariant $I(x)$ is a property attached to each point x that is preserved by the flow, $I(x) = I(f^t(x))$; for example, at point $x(t) = f^t(x)$ a green particle (more formally: a *passive scalar*) is embedded into the fluid. As $I(x)$ is invariant, its total time derivative vanishes, $\dot{I}(x) = 0$. Written in terms of partial derivatives this is the *conservation equation* for the material invariant

$$\partial_t I + v \cdot \partial I = 0. \quad (33.1)$$

Let the *density* of representative points be $\rho(x, t)$. The manner in which the flow redistributes $I(x)$ is governed by a partial differential equation whose form is relatively simple because the representative points are neither created nor destroyed. This conservation property is expressed in the integral statement

$$\partial_t \int_V dx \rho I = - \int_{\partial V} d\sigma \hat{n}_i v_i \rho I,$$

where V is an arbitrary volume in the state space $\mathcal{M}$, ∂V is its surface, $\hat{n}$ is its outward normal, and repeated indices are summed over throughout. The divergence theorem turns the surface integral into a volume integral,

$$\int_V [\partial_t(\rho I) + \partial_i(v_i \rho I)] dx = 0,$$

where ∂_i is the partial derivative operator with respect to x_i . Since the integration is over an arbitrary volume, we conclude that

$$\partial_t(\rho I) + \partial_i(\rho I v_i) = 0. \quad (33.2)$$

The choice $I \equiv 1$ yields the *continuity equation* for the density:

$$\partial_t \rho + \partial_i(\rho v_i) = 0. \quad (33.3)$$

Here we have used the language of fluid mechanics to ease the visualization, but, as we have seen in our previous derivation of the continuity equation (19.23), any deterministic state space flow satisfies the continuity equation in any dimension.

Why -even though the dynamics is nonlinear- is this equation linear? As each deterministic orbit is distinct and intersects no other orbit, no ‘particles’ are created or destroyed, they are non-interacting, hence description in terms of linear evolution operators possible.

33.2 Brownian diffusion

Consider tracer molecules, let us say big, laggardly green molecules, embedded in a denser gas of light molecules. Assume that the density of tracer molecules $\bar{\rho}$ compared to the background gas density ρ is low, so we can neglect green-green collisions. Each green molecule, jostled by frequent collisions with the background gas, executes its own Brownian motion. The molecules are neither created nor destroyed, so their number within an arbitrary volume V changes with time only by the current density $\bar{j}_i$ flow through its surface ∂V (with $\hat{n}$ its outward normal):

$$\partial_t \int_V dx \bar{\rho} = - \int_{\partial V} d\sigma \hat{n}_i \bar{j}_i. \quad (33.4)$$

The divergence theorem turns this into the conservation law for tracer density:

$$\partial_t \bar{\rho} + \partial_i \bar{j}_i = 0. \quad (33.5)$$

The tracer density $\bar{\rho}$ is defined as the average density of a ‘material particle,’ averaged over a subvolume large enough to contain many green (and still many more background) molecules, but small compared to the macroscopic observational scales. What is $\bar{j}$? If the density is constant, on the average as many molecules leave the material particle volume as they enter it, so a reasonable phenomenological assumption is that the *average* current density (*not* the individual particle current density ρv_i in (33.3)) is driven by the density gradient

$$\bar{j}_i = -D \frac{\partial \bar{\rho}}{\partial x_i}. \quad (33.6)$$

This is the *Fick law*, with the diffusion constant D a phenomenological parameter. Substituting this current into (33.5) yields the *diffusion* or *heat* equation,

$$\frac{\partial}{\partial t} \rho(x, t) = D \frac{\partial^2}{\partial x^2} \rho(x, t) \quad (33.7)$$

(from here on we drop bar’s over $\bar{\rho}$, $\bar{j}$.) For sake of streamlining the argument we have assumed in (33.6) that diffusion in d dimensions is time-invariant, homogenous and isotropic, $\Delta(x, t) = 2D \mathbf{1}$. More generally, diffusion is described by a state space- (Toto, remember? we’re not necessarily in 3 spatial dimensions anymore) and time-dependent symmetric *diffusion tensor* $\Delta_{ij} = \Delta_{ji}$, with $j_i = -\frac{1}{2} \Delta_{ij} \partial_j \rho$, leading to the anisotropic diffusion equation

$$\partial_t \rho(x, t) = \frac{1}{2} \partial_i \left(\Delta_{ij}(x, t) \partial_j \rho(x, t) \right). \quad (33.8)$$

In practice, the diffusion tensor is almost always anisotropic. For example, physicist's Brownian diffusion is a flow in the 6-dimensional {configuration, velocity} phase space, with white noise probability distribution $\exp(-\mathbf{v}^2/2k_B T)$, modeling random force kicks applied only to the 3 velocity variables $\mathbf{v}$. In this case one thinks of diffusion coefficient $D = k_B T/2$ as temperature.

33.2.1 Heat kernel

The heat equation we can solve analytically. Fourier transforming (33.7),

$$\frac{\partial}{\partial t} \tilde{\rho}(k, t) = -D k^2 \tilde{\rho}(k, t), \quad \rho(x, t) = \int \frac{dk}{2\pi} \tilde{\rho}(k, t) e^{ikx}, \quad (33.9)$$

substituting $\tilde{\rho}$, integrating over time,

$$\rho(x, t) = \int \frac{dk}{2\pi} \tilde{\rho}(k, 0) e^{ikx - D k^2 t},$$

and Fourier transforming back we obtain an exact solution of the heat equation in terms of an initial Dirac delta density distribution, $\rho(x, 0) = \delta(x - x_0)$,

$$\rho(x, t) = \frac{1}{(4\pi D t)^{d/2}} e^{-\frac{(x-x_0)^2}{4Dt}}. \quad (33.10)$$

A field theorist will see this as reminiscent of the quantum free particle propagation (see sect. 38.2.2). The solution of the general, anisotropic case (33.8) (here exact for all times) is

$$\mathcal{L}_{FP}(x, t; x_0, 0) = \frac{1}{\sqrt{\det(2\pi t \Delta)}} \exp\left[-\frac{1}{2t} (x^\top \frac{1}{\Delta} x)\right]. \quad (33.11)$$

The average distance covered in time t obeys the diffusion formula

$$\langle (x - x_0)^2 \rangle_t = \int dx \rho(x, t) (x - x_0)^2 = 2dDt. \quad (33.12)$$

The Einstein Brownian diffusion formula describes the 3-dimensional Brownian motion of heavy molecules jostled by thermal motions of the light molecules of the fluid. Here the diffusion is something much more general: it takes place in the dynamical state space of dimension d , and is described by diffusion tensor $D \rightarrow \Delta_{ij}(x, t)$.

33.2.2 Random walks

More insight into the evolution of the density of tracer molecules is obtained considering a d -dimensional random walk of an individual tracer molecule kicked by a stochastic term,

$$\frac{dx}{dt} = \hat{\xi}(t). \quad (33.13)$$

A way to make sense of $\hat{\xi}(t)$ is to first construct the probability distribution for additive noise ξ at short but finite time steps $\delta\tau$, with $t_{n+1} = t_n + \delta\tau$, and the particle $x_n = x(t_n)$ at time t_n executing a random walk, $x_{n+1} = x_n + \xi(t_n)$, where x is a d -dimensional state vector, and $x_{n,j}$ is its j th component at time n . The natural choice is that probability that the trajectory reaches x_{n+1} is given by a normalized Gaussian

$$\mathcal{L}_{FP}(x_{n+1}, t_{n+1}; x_n, t_n) = \frac{1}{\sqrt{\det(2\pi\delta\tau\Delta)}} \exp\left[-\frac{1}{2\delta\tau}(\xi_n^\top \frac{1}{\Delta}\xi_n)\right], \quad (33.14)$$

$\xi_n = x_{n+1} - x_n$, characterized by zero mean and the diffusion tensor (covariance matrix),

$$\langle \xi_j(t_n) \rangle = 0, \quad \langle \xi_i(t_m) \xi_j^\top(t_n) \rangle = \delta\tau \Delta_{ij} \delta_{nm}, \quad (33.15)$$

where $\langle \cdots \rangle$ stands for ensemble average over many realizations of the noise, and the superfix $^\top$ indicates a transpose. As the time discretization $\delta\tau$ is arbitrary, the diffusing cloud of noisy trajectories should be described by a distribution that keeps its form as $\delta\tau \rightarrow 0$. Indeed, the semigroup property of a Gaussian kernel,

$$\mathcal{L}_{FP}(x, t; x'', t'') = \int dx' \mathcal{L}_{FP}(x, t; x', t') \mathcal{L}_{FP}(x', t'; x'', t''), \quad (33.16)$$

ensures that the distribution keeps its form under successive diffusive steps. $\mathcal{L}_{FP}(x, t; x_0, 0)$ describes the diffusion at any time, including the integer time increments $\{t_n\} = \{\delta\tau, 2\delta\tau, \dots, n\delta\tau, \dots\}$, and thus provides a bridge between the continuous and discrete time formulations of noisy evolution.



example 33.1
p. 689

33.3 Noisy trajectories

The connection between path integration and Brownian motion is so close that they are nearly indistinguishable. Unfortunately though, like a body and its mirror image, the sum over paths for Brownian motion is a theory having substance, while its path integral image exists mainly in the eye of the beholder.

—L. S. Schulman

(P. Cvitanović and D. Lippolis)

So far we have considered tracer molecule dynamics which is purely Brownian, with no deterministic ‘drift’. Consider next a d -dimensional deterministic flow $\dot{x} = v(x)$ perturbed by a stochastic term $\hat{\xi}$,

$$\frac{dx}{dt} = v(x) + \hat{\xi}(t), \quad (33.17)$$

where the deterministic velocity field $v(x)$ is called ‘drift’ in the stochastic literature (or ‘advection’ in fluid dynamics), and $\hat{\xi}(t)$ is additive noise, uncorrelated in

time. We shall refer to equations of this type as *Langevin equations*. The more general case of a tensor $\Delta(x)$ which is a state space position dependent but time independent can be treated along the same lines. In this case the stochastic flow (33.17) is written as

$$dx = v(x) dt + \sigma(x) d\hat{\xi}(t), \quad \langle \xi_n \xi_m^\top \rangle = \mathbf{1} \delta_{nm}, \quad \Delta = \sigma \sigma^\top. \quad (33.18)$$

$\sigma(x)$ is called the ‘diffusion matrix’, and the noise is referred to as ‘multiplicative’. Explicit time dependence in $\Delta(x, t)$ would take us into world of non-autonomous, externally driven flows, beyond the comfort zone of ChaosBook.org.

As in (33.13), a way to make sense of (33.17) is to first construct the probability distribution for additive noise ξ at a short but finite time $\delta\tau$. In time $\delta\tau$ the deterministic trajectory advances by $v(x_n) \delta\tau$. As $\delta\tau$ is arbitrary, it is desirable that the diffusing cloud of noisy trajectories is given by a distribution that keeps its form as $\delta\tau \rightarrow 0$. This holds if the noise is Brownian, i.e., the probability that the trajectory reaches x_{n+1} is given by a normalized Gaussian (33.14),

$$\mathcal{L}_{FP}(x_{n+1}, \delta\tau; x_n, 0) = \frac{1}{N} \exp \left[-\frac{1}{2\delta\tau} (\xi_n^\top \frac{1}{\Delta_n} \xi_n) \right]. \quad (33.19)$$

Here $\xi_n = \delta x_n - v(x_n) \delta\tau$, the deviation of the noisy trajectory from the deterministic one, can be viewed either in terms of velocities $\{\dot{x}, v(x)\}$ (continuous time formulation), or finite time maps $\{x_n \rightarrow x_{n+1}, x_n \rightarrow f^{\delta\tau}(x_n)\}$ (discrete time formulation),

$$\delta x_n = x_{n+1} - x_n \simeq \dot{x}_n \delta\tau, \quad f^{\delta\tau}(x_n) - x_n \simeq v(x_n) \delta\tau, \quad (33.20)$$

where

$$\{x_0, x_1, \dots, x_n, \dots, x_k\} = \{x(0), x(\delta\tau), \dots, x(n\delta\tau), \dots, x(t)\} \quad (33.21)$$

is a sequence of $k+1$ points $x_n = x(t_n)$ along the noisy trajectory, separated by time increments $\delta\tau = t/k$.

The phenomenological Fick law current (33.6) is now a sum of two components, the material particle deterministic drift $v(x)$ and the weak noise term

$$j_i = v_i \rho - D \frac{\partial \rho}{\partial x_i}, \quad \left[= v_i \rho - \frac{1}{2} \Delta_{ij}(x) \partial_j \rho(x, t) \right], \quad (33.22)$$

with the full, anisotropic and space-dependent version indicated in $[\dots]$, and sum over repeated index j implied. Substituting this j into (33.5) yields

$$\partial_t \rho + \partial_i (\rho v_i) = D \partial^2 \rho. \quad (33.23)$$

In the general, anisotropic case (33.8) this equation is known as the *Fokker-Planck* or *forward Kolmogorov* equation

$$\partial_t \rho + \partial_i (\rho v_i) = \frac{1}{2} \partial_i \left(\Delta_{ij}(x) \partial_j \rho(x, t) \right). \quad (33.24)$$

The left hand side, $d\rho/dt = \partial_t \rho + \partial \cdot (\rho v)$, is deterministic, with the continuity equation (33.3) recovered in the weak noise limit $D \rightarrow 0$. The right hand side

describes the diffusive transport in or out of the material particle volume. If the density is lower than in the immediate neighborhood, the local curvature is positive, $\partial^2 \rho > 0$, and the density grows. Conversely, for negative curvature diffusion lowers the local density, thus smoothing the variability of ρ . Where is the density going globally?

If the system is bound, the probability density vanishes sufficiently fast outside the central region, $\rho(x, t) \rightarrow 0$ as $|x| \rightarrow \infty$, and the total probability is conserved

$$\int dx \rho(x, t) = 1.$$

Any initial density $\rho(x, 0)$ is smoothed by diffusion and with time tends to the natural measure, the invariant density

$$\rho_0(x) = \lim_{t \rightarrow \infty} \rho(x, t), \quad (33.25)$$

an eigenfunction $\rho(x, t) = e^{st} \rho_0(x)$ of the time-independent Fokker-Planck equation

$$(\partial_i v_i - D \partial^2 + s_\alpha) \rho_\alpha = 0, \quad (33.26)$$

with vanishing eigenvalue $s_0 = 0$. Provided the noiseless deterministic flow is hyperbolic, in the vanishing noise limit the leading eigenfunction of the Fokker-Planck equation tends to natural measure (19.15) of the corresponding deterministic flow, the leading eigenvector of the Perron-Frobenius operator.

If the system is open, there is a continuous outflow of probability from the region under study, the leading eigenvalue is contracting, $s_0 < 0$, and the density of the system tends to zero. In this case the leading eigenvalue s_0 of the time-independent Fokker-Planck equation (33.26) can be interpreted by saying that a finite density can be maintained by pumping back probability into the system at a constant rate $\gamma = -s_0$. The value of γ for which any initial probability density converges to a finite stationary equilibrium density is called the *escape rate*. In the noiseless limit this coincides with the deterministic escape rate (20.25).

The distribution (33.14) describes how an initial density of particles concentrated in a Dirac delta function at x_n spreads in time $\delta\tau$. In the Fokker-Planck description individual noisy Langevin trajectories (33.17) are replaced by the evolution of the density of noisy trajectories. The finite time Fokker-Planck evolution $\rho(x, t) = [\mathcal{L}_{FP}^t \circ \rho](x, 0)$ of an initial density $\rho(x_0, 0)$ is obtained by a sequence of consecutive short-time steps (33.14)

$$\mathcal{L}_{FP}(x_k, t; x_0, 0) = \int [dx] \exp \left\{ -\frac{1}{4D\delta\tau} \sum_{n=0}^{k-1} [x_{n+1} - f^{\delta\tau}(x_n)]^2 \right\}, \quad (33.27)$$

where $t = k\delta\tau$, and the Gaussian normalization factor in (33.14) is absorbed into intermediate integrations by defining

$$\begin{aligned} [dx] &= \prod_{n=0}^{k-1} \frac{dx_n^d}{N_n} \\ N_n &= [\det(2\pi\delta\tau\Delta(x_n))]^{1/2} \quad (\text{anisotropic diffusion tensor } \Delta) \\ &= (4D\delta\tau)^{d/2} \quad (\text{isotropic diffusion, } \Delta(x) = 2D\mathbf{1}). \end{aligned} \quad (33.28)$$

As $D \rightarrow 0$, the distribution tends to the noiseless, deterministic Dirac delta function Perron-Frobenius operator (19.10). The stochastic flow (33.17) can now be understood as the continuous time, $\delta\tau \rightarrow 0$ limit, with the velocity noise $\hat{\xi}(t)$ a Gaussian random variable of zero mean and covariance matrix

$$\langle \hat{\xi}_j(t) \rangle = 0, \quad \langle \hat{\xi}_i(t) \hat{\xi}_j(t') \rangle = \Delta_{ij} \delta(t - t'). \quad (33.29)$$

It is worth noting that the continuous time flow noise $\hat{\xi}(t)$ in (33.17) and (33.29) is dimensionally a velocity $[x]/[t]$, as $\mathcal{L}_{FP}(x_{n+1}, \delta\tau; x_n, 0)$ is a probability density for velocity ξ , while the discrete time noise ξ_n in (33.14), (33.15) is dimensionally a length $[x]$, as $\rho(x, t)$ is a probability density for position x . The important point is that the same diffusion tensor $\Delta(x)$ describes the diffusion both in the configuration space and the velocity space.

The continuous time limit of (33.27), $\delta\tau = t/k \rightarrow 0$, defines formally the Fokker-Planck evolution operator

$$\mathcal{L}_{FP}(x, t; x_0, 0) = \int [dx] \exp \left\{ -\frac{1}{4D} \int_0^t [\dot{x}(\tau) - v(x(\tau))]^2 d\tau \right\} \quad (33.30)$$

as a stochastic path (or Wiener) integral for a noisy flow, and the associated continuous time Fokker-Planck (or forward Kolmogorov) equation (33.24) describes the time evolution of a density of noisy trajectories. We have introduced noise phenomenologically, and used the weak noise assumption in retaining only the first derivative of ρ in formulating the Fick law (33.6) and including noise additively in (33.22). The $\delta\tau \rightarrow 0$ limit and the proper definition of $\dot{x}(\tau)$ are delicate issues of no import for the applications studied here. A full theory of stochastic ODEs is much subtler, but this will do for our purposes.

The exponent

$$-\frac{1}{4D\delta\tau} [x_{n+1} - f^{\delta\tau}(x_n)]^2 \simeq -\frac{\delta\tau}{4D} [\dot{x}(\tau) - v(x(\tau))]^2 \quad (33.31)$$

can be interpreted as a cost function which penalizes deviation of the noisy trajectory δx from its deterministic prediction $v\delta\tau$, or, in the continuous time limit, the deviation of the noisy trajectory tangent $\dot{x}$ from the deterministic velocity field v . Its minimization is one of the most important tools of the optimal control theory, with velocity $\dot{x}(\tau)$ along a trial path varied with aim of minimizing its distance to the target $v(x(\tau))$.

33.4 Noisy maps

(P. Cvitanović and D. Lippolis)

For pedagogical reasons we shall often find it convenient to consider a noisy *discrete time* dynamical system

$$x_{n+1} = f(x_n) + \xi_n, \quad (33.32)$$

where x is a d -dimensional state vector, and $x_{n,j}$ is its j th component at time n . In the Fokker-Planck description individual noisy trajectories are replaced by the evolution of the density of noisy trajectories, with the $\xi_n = x_{n+1} - f(x_n)$ probability distribution of zero mean and diffusion tensor, and the time increment in (33.15) set to $\delta\tau = 1$,

$$\langle \xi_{n,j} \rangle = 0, \quad \langle \xi_{n,i} \xi_{m,j}^\top \rangle = \Delta_{ij}(x_n) \delta_{nm}. \quad (33.33)$$

As we shall show, in nonlinear dynamics the noise is *never* isotropic and/or homogeneous. Even if the infinitesimal time step noise (33.14) covariance matrix in (33.18) were independent of the state space position x , this cannot be true of $\Delta(x)$ for the discrete time flow (33.32) obtained by the Poincaré section reduction method of sect. 3.1, as the return times (3.1) and the noise accumulated along the corresponding trajectory segments depend on the starting Poincaré section point. Indeed, as we shall argue in sect. 33.5, in nonlinear dynamics *all noise is local*. As long as the noise distribution at x is autonomous (not explicitly dependent on time) the stochastic flow (33.32) can be written as $x_{n+1} = x_n + \sigma(x_n) \xi_n$, where $\Delta = \sigma \sigma^\top$, and $\sigma(x)$ is the multiplicative noise diffusion matrix defined in (33.18).

The action of discrete one-time step *Fokker-Planck evolution operator* on the density distribution ρ at time k ,

$$\begin{aligned} \rho_{k+1}(y) &= [\mathcal{L}_{FP} \rho_k](y) = \int dx \mathcal{L}_{FP}(y, x) \rho_k(x) \\ \mathcal{L}_{FP}(y, x) &= \frac{1}{N(x)} e^{-\frac{1}{2}(y-f(x))^\top \frac{1}{\Delta(x)}(y-f(x))}, \end{aligned} \quad (33.34)$$

is centered on the deterministic step $f(x)$ and smeared out diffusively by noise. Were diffusion uniform and isotropic, $\Delta(x) = 2D\mathbf{1}$, the Fokker-Planck evolution operator would be proportional to $\exp(-\{y - f(x)\}^2/2\Delta)$, i.e., the penalty for straying from the deterministic path is just a quadratic error function. The k th iterate of $\mathcal{L}_{FP}^k(x_k; x_0) = \mathcal{L}_{FP}(x_k, t; x_0, 0)$ is a d -dimensional path integral over the $k - 1$ intermediate noisy trajectory points,

$$\mathcal{L}_{FP}^k(x_k; x_0) = \int [dx] e^{-\frac{1}{2} \sum_n (x_{n+1} - f(x_n))^\top \frac{1}{\Delta(x_n)} (x_{n+1} - f(x_n))}, \quad (33.35)$$

where the Gaussian normalization factor in (33.34) is absorbed into intermediate integrations by defining

$$[dx] = \prod_{n=1}^{k-1} \frac{dx_n^d}{N_n}, \quad N_n = \sqrt{(2\pi)^d \det \Delta(x_n)}. \quad (33.36)$$

We shall also need to determine the effect of noise accumulated along the trajectory points *preceding* x . As the noise is additive forward in time, one cannot simply invert the Fokker-Planck evolution operator; instead, the past is described by the *adjoint Fokker-Planck evolution operator*,

$$\tilde{\rho}_{k-1}(x) = [\mathcal{L}_{FP}^\dagger \tilde{\rho}_k](x) = \int [dy] e^{-\frac{1}{2}(y-f(x))^\top \frac{1}{\Delta}(y-f(x))} \tilde{\rho}_k(y), \quad (33.37)$$

which transports a density concentrated around the point $f(x)$ to a density concentrated around the previous point x and adds noise to it. In the deterministic, vanishing noise limit this is the Koopman operator (31.1).

The Fokker-Planck evolution operator (33.34) is non-hermitian and non-unitary. For example, if the deterministic flow is contracting, the natural measure (the leading right eigenvector of the evolution operator) will be concentrated and peaked, but then the corresponding left eigenvector has to be broad and flat, as backward in time the deterministic flow is expanding. We shall denote by ρ_α the right eigenvectors of $\mathcal{L}_{FP}$, and by $\tilde{\rho}_\alpha$ its left eigenvectors, i.e., the right eigenvectors of the adjoint operator $\mathcal{L}_{FP}^\dagger$.

33.5 All nonlinear noise is local

I ain't gonna work for Maggie's pa no more
 No, I ain't gonna work for Maggie's pa no more
 Well, he puts his cigar
 Out in your face just for kicks

— Bob Dylan, *Maggie's Farm*

(P. Cvitanović and D. Lippolis)

Our main goal in this section is to convince the reader that the diffusive dynamics of nonlinear flows is *fundamentally different from Brownian motion*, with the flow inducing a local, history dependent noise. In order to accomplish this, we generalize here the notion of invariant deterministic recurrent solutions, such as fixed points and periodic orbits, to noisy flows. While a Langevin trajectory (33.32) can never return exactly to the initial point and thus cannot ever be periodic, in the Fokker-Planck formulation (33.35) a recurrent motion can be defined as one where a peaked distribution returns to the initial neighborhood after time n . Recurrence so defined not only coincides with the deterministic notion of a recurrent orbit in the vanishing noise limit, but it also enables us to derive exact formulas for how this local, history dependent noise is to be computed.

As the function $x_{n+1} - f(x_n)$ is a nonlinear function, in general the path integral (33.35) can only be evaluated numerically. In the vanishing noise limit the Gaussian kernel sharpens into the Dirac δ -function, and the Fokker-Planck evolution operator reduces to the deterministic Perron-Frobenius operator (19.10). For weak noise the Fokker-Planck evolution operator can be evaluated perturbatively as an asymptotic series in powers of the diffusion constant, centered on the deterministic trajectory. Here we retain only the linear term in this series, which has a particularly simple dynamics given by a covariance matrix evolution formula (see (33.45) below) that we now derive.

We shift local coordinates labeled at time ' a ' to the deterministic trajectory $\{\dots, x_{-1}, x_0, x_1, x_2, \dots\}$ centered coordinate frame $x = x_a + z_a$, Taylor expand $f(x) = f_a(z_a) = x_{a+1} + M_a z_a + \dots$, and approximate the noisy map (33.32) by its

linearization,

$$z_{a+1} = M_a z_a + \xi_a, \quad M_{ij}(x) = \partial f_i / \partial x_j, \quad (33.38)$$

with the deterministic trajectory points at $z_a = z_{a+1} = 0$, and $M_a = M(x_a)$ the one time step Jacobian matrix. The corresponding linearized Fokker-Planck evolution operator (33.34) action on density $\rho_a(z_a) = \rho(x_a + z_a, a)$ is given in the local coordinates by

$$\rho_{a+1}(z_{a+1}) = \int dz_a \mathcal{L}_{FP}^a(z_{a+1}, z_a) \rho_a(z_a) \quad (33.39)$$

by the linearization (33.38) centered on the deterministic trajectory

$$\mathcal{L}_{FP}^a(z_{a+1}, z_a) = \frac{1}{N} e^{-\frac{1}{2}(z_{a+1} - M_a z_a)^\top \frac{1}{\Delta_a} (z_{a+1} - M_a z_a)}. \quad (33.40)$$

The superscript ‘ a ’ in $\mathcal{L}_{FP}^a$ distinguishes the local, linearized Fokker-Planck evolution operator coordinate frame $z_a = x - x_a$ centered on the deterministic trajectory point x_a from the full global evolution operator (33.35), in global coordinate system x .

The kernel of the linearized Fokker-Planck evolution operator (33.40) is a Gaussian. As a convolution of a Gaussian with a Gaussian is again a Gaussian, we investigate the action of the linearized Fokker-Planck evolution operator on a normalized, cigar-shaped Gaussian density distribution

$$\rho_a(z) = \frac{1}{C_a} e^{-\frac{1}{2} z^\top \frac{1}{Q_a} z}, \quad C_a = [\det(2\pi Q_a)]^{1/2}, \quad (33.41)$$

and the action of the linearized adjoint Fokker-Planck evolution operator on density

$$\tilde{\rho}_a(z) = \frac{1}{C_a} e^{-\frac{1}{2} z^\top \frac{1}{\tilde{Q}_a} z}, \quad C_a = [\det(2\pi \tilde{Q}_a)]^{1/2}, \quad (33.42)$$

also centered on the deterministic trajectory, with strictly positive $[d \times d]$ covariance matrices $Q, \tilde{Q}$. Label ‘ a ’ plays a double role, and $\{a+1, a\}$ stands both for the {next, initial} space partition and for the times the trajectory lands in these partitions. The linearized Fokker-Planck evolution operator (33.40) maps the Gaussian $\rho_a(z_a)$ into the Gaussian

$$\rho_{a+1}(z_{a+1}) = \frac{1}{C_a} \int [dz_a] e^{-\frac{1}{2} [(z_{a+1} - M_a z_a)^\top \frac{1}{\Delta_a} (z_{a+1} - M_a z_a) + z_a^\top \frac{1}{Q_a} z_a]} \quad (33.43)$$

one time step later. Likewise, linearizing the adjoint Fokker-Planck evolution operator (33.37) around the x_a trajectory point yields:

$$\tilde{\rho}_a(z_a) = \frac{1}{C_{a+1}} \int [dz_{a+1}] e^{-\frac{1}{2} [(z_{a+1} - M_a z_a)^\top \frac{1}{\Delta_a} (z_{a+1} - M_a z_a) + z_{a+1}^\top \frac{1}{\tilde{Q}_{a+1}} z_{a+1}]} \quad (33.44)$$

Completing the squares, integrating and substituting (33.41), respectively (33.42) we obtain the formula for covariance matrix evolution forward in time,

$$Q_{a+1} = M_a Q_a M_a^\top + \Delta_a. \quad (33.45)$$

In the adjoint case, the evolution of the $\tilde{Q}$ is given by

$$M_a \tilde{Q}_a M_a^\top = \tilde{Q}_{a+1} + \Delta_a. \quad (33.46)$$

The two covariance matrices differ, as the adjoint evolution $\tilde{Q}_a$ is computed by going backwards along the trajectory. These *covariance evolution* rules are the basis of all that follows.

Think of the initial covariance matrix (33.41) as an error correlation matrix describing the precision of the initial state, a cigar-shaped probability distribution $\rho_a(z_a)$. In one time step this density is deterministically advected and deformed into density with covariance $MQM^\top$, and then the noise Δ is added: the two kinds of independent uncertainties add up as sums of squares, hence the covariance evolution law (33.45), resulting in the Gaussian ellipsoid whose widths and orientation are given by the singular values and singular vectors (6.15) of the covariance matrix. After n time steps, the variance Q_a is built up from the deterministically propagated $M_a^n Q_{a-n} M_a^{nT}$ initial distribution, and the sum of noise kicks at intervening times, $M_a^k \Delta_{a-k} M_a^{kT}$, also propagated deterministically.

The pleasant surprise is that the evaluation of this noise requires no Fokker-Planck PDE formalism. The width of a Gaussian packet centered on a trajectory is fully specified by a deterministic computation that is already a pre-computed byproduct of the periodic orbit computations; the deterministic orbit and its linear stability. We have attached label ‘ a ’ to $\Delta_a = \Delta(x_a)$ in (33.45) to account for the noise distributions that are inhomogeneous, state space dependent, but time independent multiplicative noise.

33.6 Weak noise: Hamiltonian formulation

All imperfection is easier to tolerate if served up in small doses.

— Wislawa Szymborska

(G. Vattay and P. Cvitanović)

In the spirit of the WKB approximation (to be fully developed in chapter 37), we shall now study the evolution of the probability distribution by rewriting it as

$$\rho(x, t) = e^{\frac{1}{2D} R(x, t)}. \quad (33.47)$$

The time evolution of R is given by

$$\partial_t R + v \partial R + (\partial R)^2 = D \partial^2 v + D \partial^2 R.$$

Consider now the weak noise limit and drop the terms proportional to D . The remaining equation

$$\partial_t R + H(x, \partial R) = 0$$

is known as the Hamilton-Jacobi equation. The function R can be interpreted as the Hamilton's principal function, corresponding to the Hamiltonian section 38.1

$$H(x, p) = p v(x) + p^2/2,$$

with the Hamilton's equations of motion

$$\begin{aligned}\dot{x} &= \partial_p H = v + p \\ \dot{p} &= -\partial_x H = -A^\top p,\end{aligned}\tag{33.48}$$

where A is the stability matrix (4.3)

$$A_{ij}(x) = \frac{\partial v_i(x)}{\partial x_j}.$$

The noise Lagrangian is then

$$L(x, \dot{x}) = \dot{x} \cdot p - H = \frac{1}{2} [\dot{x} - v(x)]^2.\tag{33.49}$$

We have come the full circle - the Lagrangian is the exponent of our assumed Gaussian distribution (33.31) for noise $\xi^2 = [\dot{x} - v(x)]^2$. What is the meaning of this Hamiltonian, Lagrangian? Consider two points x_0 and x . Which noisy path is the most probable path that connects them in time t ? The probability of a given path $\mathcal{P}$ is given by the probability of the noise sequence $\xi(t)$ which generates the path. This probability is proportional to the product of the noise probability functions (33.31) along the path, and the total probability for reaching x from x_0 in time t is given by the sum over all paths, or the stochastic path integral (Wiener integral)

$$\begin{aligned}P(x, x_0, t) &\sim \sum_{\mathcal{P}} \prod_j p(\xi(\tau_j), \delta\tau_j) = \int \prod_j d\xi_j \left(\frac{\delta\tau_j}{4\pi D} \right)^{d/2} e^{-\frac{\xi(\tau_j)^2}{4D} \delta\tau_j} \\ &\rightarrow \frac{1}{Z} \sum_{\mathcal{P}} \exp\left(-\frac{1}{4D} \int_0^t d\tau \xi^2(\tau)\right),\end{aligned}\tag{33.50}$$

where $\delta\tau_i = \tau_i - \tau_{i-1}$, and the normalization constant is

$$\frac{1}{Z} = \lim \prod_i \left(\frac{\delta\tau_i}{2\pi D} \right)^{d/2}.$$

The most probable path is the one maximizing the integral inside the exponential. If we express the noise (33.17) as

$$\xi(t) = \dot{x}(t) - v(x(t)),$$

the probability is maximized by the variational principle

$$\min \int_0^t d\tau [\dot{x}(\tau) - v(x(\tau))]^2 = \min \int_0^t L(x(\tau), \dot{x}(\tau)) d\tau.$$

By the standard arguments, for a given x , x' and t the probability is maximized by a solution of Hamilton's equations (33.48) that connects the two points $x_0 \rightarrow x'$ in time t . The solution is a bit boring: $\dot{x} = v$, $p = 0$, and lives in the initial, d -dimensional state space, so not much is to be made of this surprising appearance of Hamiltonians.

Résumé

When a deterministic trajectory is smeared out under the influence of Gaussian noise of strength D , the deterministic dynamics is recovered in the weak noise limit $D \rightarrow 0$. The effect of the noise can be taken into account by adding noise corrections to the deterministic trace formula.

Symplectic structure emerges here not as a deep principle of mechanics, but an artifact of the leading approximation to quantum/noisy dynamics, not respected by higher order corrections. The same is true of semiclassical quantum dynamics; higher corrections do not respect canonical invariance.

Commentary

Remark 33.1. A brief history of noise. The theory of stochastic processes is a vast subject, starting with the Laplace 1810 memoir [29], spanning over centuries, and over disciplines ranging from pure mathematics to impure finance. The presentation given here is based on the Cvitanović and Lippolis 2012 Maribor lectures [11]. The material reviewed is standard [2, 26, 36], but needed in order to set the notation for what is new here, the role that local Fokker-Planck operators play in defining stochastic neighborhoods of periodic orbits. We enjoyed reading van Kampen classic [26], especially his railings against those who blunder carelessly into nonlinear landscapes. Having committed this careless chapter to print, we shall no doubt be cast to a special place on the long list of van Kampen's sinners (and not for the first time, either). A more specialized monograph like Risken's [36] will do just as well. Schulman's monograph [38] contains a very readable summary of Kac's [24] exposition of Wiener's integral over stochastic paths. The standard Langevin equation [7] is a stochastic equation for a Brownian particle, in which one replaces the Newton's equation for force by two counter-balancing forces: random accelerations $\xi(t)$ which tend to smear out a particle trajectory, and a damping term which drives the velocity to zero. In this context D is Einstein diffusion constant, and (33.12) is the Einstein diffusion formula [16]. Here we denote by 'Langevin equation' a more general family of stochastic differential equations (33.17) with additive or multiplicative [15, 28] weak noise. Noisy *discrete time* dynamical systems are discussed in refs. [6, 17, 27].

In probabilist literature [4] the differential operator $-\nabla \cdot (v(x)\rho(x, t)) + D \nabla^2 \rho(x, t)$ is called 'Fokker-Planck operator;' here we reserve the term 'Fokker-Planck evolution operator' for the finite time, 'Green's function' integral operator (33.30), i.e., the stochastic

path (Wiener) integral [31, 36, 39] for a noisy flow, with the associated continuous time Fokker-Planck [26, 32, 36] (or forward Kolmogorov) equation (33.24).

The cost function (33.31) appears to have been first introduced by Wiener as the exact solution for a purely diffusive Wiener-Lévy process in one dimension, see (33.51). Onsager and Machlup [20, 33] use it in their variational principle to study thermodynamic fluctuations in a neighborhood of single, linearly attractive equilibrium point (i.e., without any dynamics). It plays important role in the optimal control theory [3, 34]. Gaussians are often rediscovered, so Onsager-Machlup seminal paper, which studies the same attractive linear fixed point, is in literature often credited for being the first to introduce a variational method -the ‘principle of least dissipation’- based on the Lagrangian of form (33.49). They, in turn, credit Rayleigh [35] with introducing the least dissipation principle in hydrodynamics. Onsager-Machlup paper deals only with a finite set of linearly damped thermodynamic variables, and not with a nonlinear flow or unstable periodic orbits.

Gaspard [21] derives a trace formula for the Fokker-Planck equation associated with Itô stochastic differential equations describing noisy time-continuous nonlinear dynamical systems. In the weak-noise limit, the trace formula provides estimations of the eigenvalues of the Fokker-Planck operator on the basis of the Pollicott-Ruelle resonances of the noiseless deterministic system, which is assumed to be non-bifurcating. At first order in the noise amplitude, the effect of noise on a periodic orbit is given in terms of the period and the derivative of the period with respect to the pseudo-energy of the Onsager-Machlup-Freidlin-Wentzell scheme [20]. The dynamical ‘action’ Lagrangian in the exponent of (33.30), and the associated symplectic Hamiltonian were first written down in 1970’s by Freidlin and Wentzell [20], whose formulation of the ‘large deviation principle’ was inspired by the Feynman quantum path integral [18]. Feynman, in turn, followed Dirac [14] who was the first to discover that in the short-time limit the quantum propagator (imaginary time, quantum sibling of the Wiener stochastic distribution (33.51)) is exact. Gaspard [21] thus refers to the ‘pseudo-energy of the Onsager-Machlup-Freidlin-Wentzell scheme.’ M. Roncadelli [13, 37] refers to the Fokker-Planck exponent in (33.30) as the ‘Wiener-Onsager-Machlup Lagrangian,’ constructs weak noise saddle-point expansion and writes transport equations for the higher order coefficients. In our exposition the setting is more general: we study fluctuations over a state space-varying velocity field $v(x)$.

Remark 33.2. Weak noise perturbation theory. DasBuch omits any discussion of the Martin-Siggia-Rose [31] type weak noise corrections. For an overview of possible ways for improvement of diagrammatic summation in noisy field theories, see *Chaotic Field Theory: a Sketch* [8]. The details are in the three papers on trace formulas for stochastic evolution operators (see also ref. [37]): Weak noise perturbation theory [9], smooth conjugation method [10], and local matrix representation approach [12]. Such corrections have not been evaluated before, probably because one is so unsure about nature of the noise itself that n th order correction is beyond the point. Doing continuous time flows requires the same kind of corrections, with diagrams standing for integrals rather than sums, though no one ever tried weakly stochastic flows in continuous time.

Remark 33.3. Covariance evolution. In quantum mechanics the linearized evolution operator corresponding to the linearized Fokker-Planck evolution operator (33.40) is known as the Van Vleck propagator, the basic block in the semiclassical periodic orbit quantization [22], see chapter 38. Q covariance matrix composition rule (33.45) or its continuous time version is called ‘covariance evolution’ (for example, in ref. [41]), but it

goes all the way back to Lyapunov's 1892 thesis [30]. In the Kalman filter literature [1, 25] it is called 'prediction'.

Remark 33.4. Operator ordering. According to L. Arnold [2] review of the original literature, the derivations are much more delicate than what is presented here: the noise is *colored* rather than Dirac delta function in time. He refers only to the linear case as the 'Langevin equation'. The $\delta\tau \rightarrow 0$ limit and the proper definition of $\dot{x}(\tau)$ are delicate issues [2, 19, 23, 40] of no import for the applications of stochasticity studied here: Itô and Stratanovich operator ordering issues arise in the order beyond the leading approximation considered here.

References

- [1] H. D. I. Abarbanel, D. R. Creveling, R. Farsian, and M. Kostuk, "Dynamical state and parameter estimation", *SIAM J. Appl. Math.* **8**, 1341–1381 (2009).
- [2] L. Arnold, *Stochastic Differential Equations: Theory and Applications* (Wiley, New York, 1974).
- [3] R. Bellman, *Dynamic Programming* (Princeton Univ. Press, Princeton, 1957).
- [4] N. Berglund and B. Gentz, *Noise-Induced Phenomena in Slow-Fast Dynamical Systems: A Sample-Paths Approach* (Springer, Berlin, 2005).
- [5] N. Bleistein and R. A. Handelsman, *Asymptotic Expansions of Integrals* (Dover, New York, 1986).
- [6] A. Boyarsky, "On the significance of absolutely continuous invariant measures", *Physica D* **11**, 130–146 (1984).
- [7] S. Chandrasekhar, "Stochastic problems in physics and astronomy", *Rev. Mod. Phys.* **15**, 1–89 (1943).
- [8] P. Cvitanović, "Chaotic Field Theory: A sketch", *Physica A* **288**, 61–80 (2000).
- [9] P. Cvitanović, C. P. Dettmann, R. Mainieri, and G. Vattay, "Trace formulas for stochastic evolution operators: Weak noise perturbation theory", *J. Stat. Phys.* **93**, 981–999 (1998).
- [10] P. Cvitanović, C. P. Dettmann, R. Mainieri, and G. Vattay, "Trace formulae for stochastic evolution operators: Smooth conjugation method", *Nonlinearity* **12**, 939 (1999).
- [11] P. Cvitanović and D. Lippolis, Knowing when to stop: How noise frees us from determinism, in *Let's Face Chaos through Nonlinear Dynamics*, edited by M. Robnik and V. G. Romanovski (2012), pp. 82–126.
- [12] P. Cvitanović, N. Sørengaard, G. Palla, G. Vattay, and C. P. Dettmann, "Spectrum of stochastic evolution operators: Local matrix representation approach", *Phys. Rev. E* **60**, 3936–3941 (1999).

- [13] A. Defendi and M. Roncadelli, “Semiclassical approximation as a small-noise expansion”, *J. Phys. A* **28**, 515–520 (1995).
- [14] P. A. M. Dirac, “The Lagrangian in quantum mechanics”, *Phys. Z. SowjUn.* **3**, 64–72 (1933).
- [15] J. L. Doob, “The Brownian movement and stochastic equations”, *Ann. Math.* **43**, 351–369 (1942).
- [16] A. Einstein, “Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen”, *Ann. Phys.* **17**, 549–560 (1905).
- [17] M. J. Feigenbaum and B. Hasslacher, “Irrational decimations and path-integrals for external noise”, *Phys. Rev. Lett.* **49**, 605–609 (1982).
- [18] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (Dover Publications Inc., New York, 1965).
- [19] R. F. Fox, “Gaussian stochastic processes in physics”, *Phys. Rep.* **48**, 179–283 (1978).
- [20] M. I. Freidlin and A. D. Wentzel, *Random Perturbations of Dynamical Systems* (Springer, Berlin, 1998).
- [21] P. Gaspard, “Trace formula for noisy flows”, *J. Stat. Phys.* **106**, 57 (2002).
- [22] M. C. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer, New York, 1990).
- [23] K. Itô, “Stochastic integral”, *Proc. Imp. Acad. Tokyo* **20**, 519–524 (1944).
- [24] M. Kac, *Probability and Related Topics in Physical Sciences* (Amer. Math. Soc., 1959).
- [25] R. E. Kalman, “A new approach to linear filtering and prediction problems”, *J. Basic Eng.* **82**, 35–45 (1960).
- [26] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. (Elsevier, Amsterdam, 2007).
- [27] Y. I. Kifer, “On small random perturbations of some smooth dynamical systems”, *Math. USSR-Izv.* **8**, 1083–1107 (1974).
- [28] C. Kuehn, “Deterministic continuation of stochastic metastable equilibria via Lyapunov equations and ellipsoids”, *SIAM J. Sci. Comp.* **34**, 1635–1658 (2012).
- [29] P. S. Laplace, “Mémoire sur les intégrales définies et leur application aux probabilités, et spécialement à la recherche du milieu qu’il faut choisir entre les résultats des observations”, *Mem. Acad. Sci. (I), XI, Sec. V.*, 375–387 (1810).
- [30] A. Lyapunov, “Problème général de la stabilité du mouvement”, *Ann. Math. Studies* **17**, Russian original Kharkow, 1892, 531–534 (1977).
- [31] P. C. Martin, E. D. Siggia, and H. A. Rose, “Statistical mechanics of classical systems”, *Phys. Rev. A* **8**, 423–437 (1973).
- [32] B. K. Øksendal, *Stochastic Differential Equations: An Introduction with Applications* (Springer, New York, 1998).

- [33] L. Onsager and S. Machlup, “Fluctuations and irreversible processes”, *Phys. Rev.* **91**, 1505–1512 (1953).
- [34] L. S. Pontrjagin, V. Boltyanskii, R. Gamkrelidze, and E. Mishenko, *The Mathematical Theory of Optimal Processes* (Interscience Publishers, New York, 1964).
- [35] J. W. S. Rayleigh, “LXV. On the motion of a viscous fluid”, *Philos. Mag. Ser. 6* **26**, 776–786 (1913).
- [36] H. Risken and H. Haken, *The Fokker-Planck Equation* (Springer, New York, 1996).
- [37] M. Roncadelli, “Small-fluctuation expansion of the transition probability for a diffusion process”, *Phys. Rev. E* **52**, 4661 (1995).
- [38] L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley-Interscience, New York, 1981).
- [39] B. Shraiman, C. E. Wayne, and P. C. Martin, “Scaling theory for noisy period-doubling transitions to chaos”, *Phys. Rev. Lett.* **46**, 935–939 (1981).
- [40] R. L. Stratonovich, *Conditional Markov Processes and Their Application to the Theory of Control* (Elsevier, New York, 1968).
- [41] M. K. Tippett and S. E. Cohn, “Adjoint and low-rank covariance representation”, *Nonlin. Proc. Geophys.* **8**, 331–340 (2001).

33.7 Examples

Example 33.1. Random walk in one dimension The white noise $\xi_n = x_{n+1} - x_n$ for a 1-dimensional diffusion process is a normally distributed random variable, with standard normal (i.e., Gaussian) probability distribution function,

$$\mathcal{L}_{FP}(x, t; x', t') = \frac{1}{\sqrt{4\pi D(t-t')}} \exp\left[-\frac{(x-x')^2}{4D(t-t')}\right], \quad (33.51)$$

of mean 0, variance $2D(t-t')$, and standard deviation $\sqrt{2D(t-t')}$, uncorrelated in time:

$$\langle x_{n+1} - x_n \rangle = 0, \quad \langle (x_{m+1} - x_m)(x_{n+1} - x_n) \rangle = 2D \delta\tau \delta_{mn}. \quad (33.52)$$

[return to p. 675](#)

[section 33.4](#)

Exercises

- 33.1. **Who ordered $\sqrt{\pi}$?** Derive the Gaussian integral

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx e^{-\frac{x^2}{2a}} = \sqrt{a}, \quad a > 0.$$

assuming only that you know to integrate the exponential function e^{-x} . Hint: x^2 is a radius-squared of something. π is related to the area or circumference of something.

- 33.2. **d-dimensional Gaussian integrals.** Show that the Gaussian integral in d -dimensions is given by

$$\begin{aligned} Z[J] &= \int d^d x e^{-\frac{1}{2} x^T \cdot M^{-1} \cdot x + x^T \cdot J} \\ &= (2\pi)^{d/2} |\det M|^{\frac{1}{2}} e^{\frac{1}{2} J^T \cdot M \cdot J}, \end{aligned} \quad (33.53)$$

where M is a real positive definite $[d \times d]$ matrix, i.e., a matrix with strictly positive eigenvalues, x and J are d -dimensional vectors, and $(\cdots)^T$ denotes the transpose.

This integral you will see over and over in statistical mechanics and quantum field theory: it's called 'free field theory', 'Gaussian model', 'Wick expansion', etc.. This is the starting, 'propagator' term in any perturbation expansion.

Here we require that the real symmetric matrix M in the exponent is strictly positive definite, otherwise the integral is infinite. Negative eigenvalues can be accommodated by taking a contour in the complex plane [5], see exercise 33.4 *Fresnel integral*. Zero eigenvalues require stationary phase approximations that go beyond the Gaussian saddle point approximation, typically to the Airy-function type stationary points, see exercise 33.5 *Airy function for large arguments*.

- 33.3. **Convolution of Gaussians.**

(a) Show that the Fourier transform of the convolution

$$[f * g](x) = \int d^d y f(x - y)g(y)$$

corresponds to the product of the Fourier transforms

$$[f * g](x) = \frac{1}{(2\pi)^d} \int d^d k F(k)G(k)e^{-ik \cdot x}, \quad (33.54)$$

where

$$F(k) = \int \frac{d^d x}{(2\pi)^{d/2}} f(x) e^{-ik \cdot x}, \quad G(k) = \int \frac{d^d x}{(2\pi)^{d/2}} g(x) e^{-ik \cdot x}.$$

(b) Consider two normalized Gaussians

$$\begin{aligned} f(x) &= \frac{1}{N_1} e^{-\frac{1}{2} x^T \cdot \frac{1}{\Delta_1} \cdot x}, \quad N_1 = \sqrt{\det(2\pi\Delta_1)} \\ g(x) &= \frac{1}{N_2} e^{-\frac{1}{2} x^T \cdot \frac{1}{\Delta_2} \cdot x}, \quad N_2 = \sqrt{\det(2\pi\Delta_2)} \\ 1 &= \int d^d k f(x) = \int d^d k g(x). \end{aligned}$$

Evaluate their Fourier transforms

$$F(k) = \frac{1}{(2\pi)^{d/2}} e^{\frac{1}{2}k^\top \cdot \Delta_1 \cdot k}, \quad G(k) = \frac{1}{(2\pi)^{d/2}} e^{\frac{1}{2}k^\top \cdot \Delta_2 \cdot k}.$$

Show that the convolution of two normalized Gaussians is a normalized Gaussian

$$[f * g](x) = \frac{(2\pi)^{-d/2}}{\sqrt{\det(\Delta_1 + \Delta_2)}} e^{-\frac{1}{2}x^\top \cdot \frac{1}{\Delta_1 + \Delta_2} \cdot x}.$$

In other words, covariances Δ_j add up. This is the d -dimensional statement of the familiar fact that cumulative error squared is the sum of squares of individual errors. When individual errors are small, and you are adding up a sequence of them in time, you get Brownian motion. If the individual errors are small and added independently to a solution of a deterministic equation, you get Langevin and Fokker-Planck equations.

33.4. Fresnel integral.

(a) Derive the Fresnel integral

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx e^{-\frac{x^2}{2ia}} = \sqrt{ia} = |a|^{1/2} e^{i\frac{\pi}{4} \frac{a}{|a|}}.$$

Consider the contour integral $I_R = \int_{C(R)} \exp(iz^2) dz$, where $C(R)$ is the closed circular sector in the upper half-plane with boundary points 0 , R and $R \exp(i\pi/4)$. Show that $I_R = 0$ and that $\lim_{R \rightarrow \infty} \int_{C_1(R)} \exp(iz^2) dz = 0$, where $C_1(R)$ is the contour integral along the circular sector from R to $R \exp(i\pi/4)$. [Hint: use $\sin x \geq (2x/\pi)$ on $0 \leq x \leq \pi/2$.] Then, by breaking up the contour $C(R)$ into three components, deduce that

$$\lim_{R \rightarrow \infty} \left(\int_0^R e^{ix^2} dx - e^{i\pi/4} \int_0^R e^{-r^2} dr \right)$$

vanishes, and, from the real integration $\int_0^\infty \exp(-x^2) dx = \sqrt{\pi}/2$, deduce that

$$\int_0^\infty e^{ix^2} dx = e^{i\pi/4} \sqrt{\pi}/2.$$

Now rescale x by real number $a \neq 0$, and complete the derivation of the Fresnel integral.

(b) In exercise 33.2 the exponent in the d -dimensional Gaussian integrals was real, so the real symmetric matrix M in the exponent had to be strictly positive definite. However, in quantum physics one often has to evaluate the d -dimensional Fresnel integral

$$\frac{1}{(2\pi)^{d/2}} \int d^d \phi e^{-\frac{1}{2i} \phi^\top \cdot M^{-1} \cdot \phi + i \phi \cdot J},$$

with a hermitian matrix M . Evaluate it. What are conditions on its spectrum in order that the integral be well defined?

33.5. **Airy function for large arguments.** Important contributions as stationary phase points may arise from extremal points where the first non-zero term in a Taylor expansion of the phase is of third or higher order. Such situations occur, for example, at bifurcation points or in diffraction effects, (such as waves near sharp corners, waves creeping around obstacles, etc.). In such calculations, one meets Airy functions integrals of the form

$$Ai(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dy e^{i(xy - \frac{y^3}{3})}. \quad (33.55)$$

Calculate the Airy function $Ai(x)$ using the stationary phase approximation. What happens when considering the limit $x \rightarrow 0$. Estimate for which value of x the stationary phase approximation breaks down.

- 33.6. **Solving the Lyapunov differential equation.** Continuous time Lyapunov evolution equation for a covariance matrix $Q(t)$ is given by

$$\dot{Q} = A Q + Q A^\top + \Delta, \quad (33.56)$$

where $\{Q, A, \Delta\}$ are $[d \times d]$ matrices. The superscript $()^\top$ indicates the transpose of the matrix. The stability matrix $A = A(x)$ and the noise covariance matrix $\Delta = \Delta(x)$ are given. They are evaluated on a trajectory $x(t)$, and thus vary in time, $A(t) = A(x(t))$ and $\Delta(t) = \Delta(x(t))$. Determine the covariance matrix $Q(t)$ for a given initial condition $Q(0)$, by taking the following steps:

- (a) Write the solution in the form

$$Q(t) = J(t)[Q(0) + W(t)]J^\top(t),$$

with the Jacobian matrix $J(t)$ satisfying

$$\dot{J}(t) = A(t) J(t), \quad J(0) = I, \quad (33.57)$$

with I the $[d \times d]$ identity matrix. The Jacobian matrix at time t ,

$$J(t) = \hat{T} e^{\int_0^t d\tau A(\tau)}, \quad (33.58)$$

where $\hat{T}$ denotes the ‘time-ordering’ operation, can be evaluated by integrating (33.57).

- (b) Show that $W(t)$ satisfies

$$\dot{W} = \frac{1}{J} \Delta \frac{1}{J^\top}, \quad W(0) = 0. \quad (33.59)$$

- (c) Integrate (33.56) to obtain

$$Q(t) = J(t) \left[Q(0) + \int_0^t d\tau \frac{1}{J(\tau)} \Delta(\tau) \frac{1}{J^\top(\tau)} \right] J^\top(t). \quad (33.60)$$

- (d) If $A(t)$ commutes with itself throughout the interval $0 \leq \tau \leq t$, the time-ordering operation is redundant, and we have the explicit solution

$J(t, t') = \exp \int_{t'}^t d\tau A(\tau)$. Show that in this case the solution reduces to

$$\begin{aligned} Q(t) &= J(t) Q(0) J^\top(t) \\ &\quad + \int_0^t d\tau' J(t, \tau') \Delta(\tau') J(t, \tau')^\top. \end{aligned} \quad (33.61)$$

- (e) It is hard to imagine a time dependent $A(t) = A(x(t))$ that would be commuting. However, in the neighborhood of an equilibrium point x^* one can approximate the stability matrix with its time-independent linearization, $A = A(x^*)$. Show that in that case (33.58) reduces to

$$J(t, t') = e^{(t-t')A},$$

and (33.61) to what?

- 33.7. **Solving the Lyapunov differential equation.** Prove that if A is stable, the continuous Lyapunov equation

$$A Q + Q A^\top = -\Delta < 0.$$

has a solution

$$Q = \int_0^\infty dt e^{tA} \Delta e^{tA^\top}, \quad (33.62)$$

and that this solution is unique.

(P. Cvitanović)

- 33.8. **Solving the discrete Lyapunov equation.** Prove that if M is contracting, the discrete Lyapunov equation

$$Q - MQM^T = \Delta > 0$$

has a solution

$$Q = \sum_{k=0}^{\infty} M^k \Delta M^{kT}, \quad (33.63)$$

and that this solution is unique.

(P. Cvitanović)

- 33.9. **Continuous vs. discrete time Lyapunov equation.** Show that the continuous Lyapunov equation solution (suited to a Laplace transform),

$$Q = \int_0^{\infty} dt e^{tA} \Delta e^{tA^T}, \quad A < 0,$$

is equivalent to the discrete Lyapunov equation solution (suited to a Z-transform),

$$Q = \sum_{k=0}^{\infty} M^k \bar{\Delta} M^{kT}, \quad |M| < 1,$$

where

$$M = e^A, \quad \bar{\Delta} = \int_0^1 dt e^{tA} \Delta e^{tA^T}.$$

Parenthetically, often the notation does not distinguish $\bar{\Delta}$ from Δ . It should. (P. Cvitanović)

- 33.10. **Lyapunov differential equation in resolvent form.** Show that the continuous Lyapunov equation solution,

$$Q = \int_0^{\infty} dt e^{tA} \Delta e^{tA^T}, \quad A < 0,$$

is equivalent to

$$Q = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \frac{1}{i\omega - A} \Delta \frac{1}{-i\omega - A^T}.$$

(P. Cvitanović)

- 33.11. **Discrete time Lyapunov equation in the resolvent form.** Show that the continuous Lyapunov equation solution,

$$Q = \int_0^{\infty} dt e^{tA} \Delta e^{tA^T}, \quad A < 0,$$

is equivalent to the discrete Lyapunov equation solution in resolvent form,

$$Q = \frac{1}{2\pi} \int_0^{2\pi} d\omega \frac{1}{1 - e^{-i\omega} M} \bar{\Delta} \frac{1}{1 - e^{i\omega} M^T}.$$

(P. Cvitanović)

33.12. **Noise covariance matrix for a discrete time periodic orbit.**

(a) Prove that the covariance matrix at a periodic point x_a on a limit cycle p ,

$$Q_a = M_{p,a} Q_a M_{p,a}^\top + \Delta_{p,a}, \quad (33.64)$$

where

$$\begin{aligned} \Delta_{p,a} = & \Delta_a + M_{a-1} \Delta_{a-1} M_{a-1}^\top + M_{a-2}^2 \Delta_{a-2} (M_{a-2}^2)^\top \\ & + \cdots + M_{a-n_p+1}^{n_p-1} \Delta_{a-n_p+1} (M_{a-n_p+1}^{n_p-1})^\top \end{aligned} \quad (33.65)$$

is the noise accumulated per a single transversal of the periodic orbit, $M_{p,a} = M_p(x_a)$ is the cycle Jacobian matrix (4.5) evaluated on the periodic point x_a , and we have used the periodic orbit condition $x_{a+n_p} = x_a$.

(b) Derive the analogous formulas for the adjoint Fokker-Planck covariance matrix at a periodic point x_a on a repelling cycle p .

Chapter 34

Relaxation for cyclists

I cannot pass quietly over the relations between the theory of solutions of the second kind [i.e. of arbitrarily long period] and the Principle of Least Action; and it is even because of these relations that I have written chapter 29.

— H. Poincaré, Vol. 3, chap. 31, artl. 371 of *Les méthodes nouvelles de la mécanique céleste*

CYCLES, i.e., solutions of the periodic orbit condition (16.1)

$$f^{t+T}(x) = f^t(x), \quad T > 0 \quad (34.1)$$

are prerequisite to chapters 21 and 22 evaluation of spectra of deterministic evolution operators. Chapter 16 offered an introductory, hands-on guide to extraction of periodic orbits by means of the Newton-Raphson method. Here we take a very different tack, drawing inspiration from variational principles of classical mechanics, and path integrals of quantum mechanics.

In sect. 7.1.1 we converted orbits unstable forward in time into orbits stable backwards in time. Indeed, all methods for finding unstable cycles are based on the idea of constructing a new dynamical system such that (i) the position of the cycle is the same for the original system and the transformed one, (ii) the unstable cycle in the original system is a stable cycle of the transformed system.

The Newton-Raphson method for determining a fixed point x_* for a map $x' = f(x)$ is an example. The method replaces iteration of $f(x)$ by iteration of the Newton-Raphson map (7.4)

$$x'_i = g_i(x) = x_i - \left(\frac{1}{M(x) - \mathbf{1}} \right)_{ij} (f(x) - x)_j. \quad (34.2)$$

A fixed point x_* for a map $f(x)$ is also a fixed point of $g(x)$, indeed a superstable fixed point since $\partial g_i(x_*)/\partial x_j = 0$. This makes the convergence to the fixed point super-exponential.

We also learned in chapter 16 that methods that start with initial guesses for a number of points along a cycle are considerably more robust and safer than searches based on direct solution of the fixed-point condition (34.1). The relaxation (or variational) methods that we shall now describe take this multipoint approach to its logical extreme, and start by a guess of not a few points along a periodic orbit, but a guess of the entire orbit.

The idea is to make an informed rough guess of what the desired periodic orbit looks like globally, and then use variational methods to drive the initial guess toward the exact solution. Sacrificing computer memory for robustness of the method, we replace a guess that a *point* is on the periodic orbit by a guess of the *entire orbit*. And, sacrificing speed for safety, in sect. 34.1 we replace the Newton-Raphson *iteration* by a fictitious time *flow* that minimizes a cost function computed as deviation of the approximate flow from the true flow along a loop approximation to a periodic orbit.

If you have some insight into the topology of the flow and its symbolic dynamics, or have already found a set of short cycles, you might be able to construct an initial approximation to a longer cycle p as a sequence of N points $(\tilde{x}_1^{(0)}, \tilde{x}_2^{(0)}, \dots, \tilde{x}_N^{(0)})$ with the periodic boundary condition $\tilde{x}_{N+1} = \tilde{x}_1$. Suppose you have an iterative method for improving your guess; after k iterations the cost function

$$F^2(\tilde{x}^{(k)}) = \sum_i^N \left(\tilde{x}_{i+1}^{(k)} - f(\tilde{x}_i^{(k)}) \right)^2 \quad (34.3)$$

or some other more cleverly constructed function (for classical mechanics - action) is a measure of the deviation of the k th approximate cycle from the true cycle. This observation motivates variational approaches to determining cycles.

We give here three examples of such methods, two for maps, and one for billiards. In sect. 34.1 we start out by converting a problem of finding an unstable fixed point of a map into a problem of constructing a differential flow for which the desired fixed point is an attracting equilibrium point. Solving differential equations can be time intensive, so in sect. 34.2 we replace such flows by discrete iterations. In sect. 34.3 we show that for $2D$ -dimensional billiard flows variation of D coordinates (where D is the number of Hamiltonian degrees of freedom) suffices to determine cycles in the full $2D$ -dimensional phase space.

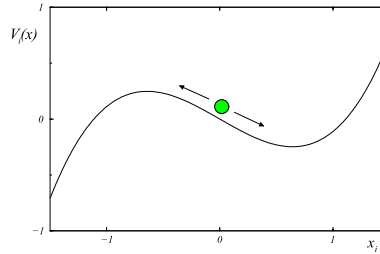
34.1 Fictitious time relaxation

(O. Biham, C. Chandre and P. Cvitanović)

The relaxation (or gradient) algorithm for finding cycles is based on the observation that a trajectory of a map such as the Hénon map (3.18),

$$\begin{aligned} x_{i+1} &= 1 - ax_i^2 + by_i \\ y_{i+1} &= x_i, \end{aligned} \quad (34.4)$$

Figure 34.1: “Potential” $V_i(x)$ (34.7) for a typical point along an initial guess trajectory. For $\sigma_i = +1$ the flow is toward the local maximum of $V_i(x)$, and for $\sigma_i = -1$ toward the local minimum. A large deviation of x_i ’s is needed to destabilize a trajectory passing through such local extremum of $V_i(x)$, hence the basin of attraction is expected to be large.



is a stationary solution of the relaxation dynamics defined by the flow

$$\frac{dx_i}{d\tau} = v_i, \quad i = 1, \dots, n \quad (34.5)$$

for any vector field $v_i = v_i(x)$ which vanishes on the trajectory. Here τ is a “fictitious time” variable, unrelated to the dynamical time (in this example, the discrete time of map iteration). As the simplest example, take v_i to be the deviation of an approximate trajectory from the exact 2-step recurrence form of the Hénon map (3.19)

$$v_i = x_{i+1} - 1 + ax_i^2 - bx_{i-1}. \quad (34.6)$$

For fixed x_{i-1} , x_{i+1} there are two values of x_i satisfying $v_i = 0$. These solutions are the two extremal points of a local “potential” function (no sum on i)

$$v_i = \frac{\partial}{\partial x_i} V_i(x), \quad V_i(x) = x_i(x_{i+1} - bx_{i-1} - 1) + \frac{a}{3}x_i^3. \quad (34.7)$$

Assuming that the two extremal points are real, one is a local minimum of $V_i(x)$ and the other is a local maximum. Now here is the idea; replace (34.5) by

$$\frac{dx_i}{d\tau} = \sigma_i v_i, \quad i = 1, \dots, n, \quad (34.8)$$

where $\sigma_i = \pm 1$.

The modified flow will be in the direction of the extremal point given by the local maximum of $V_i(x)$ if $\sigma_i = +1$ is chosen, or in the direction of the one corresponding to the local minimum if we take $\sigma_i = -1$. This is not quite what happens in solving (34.8) - all x_i and $V_i(x)$ change at each integration step - but this is the observation that motivates the method. The differential equations (34.8) then drive an approximate initial guess toward the exact trajectory. A sketch of the landscape in which x_i converges towards the proper fixed point is given in figure 34.1. As the “potential” function (34.7) is not bounded for a large $|x_i|$, the flow diverges for initial guesses which are too distant from the true trajectory. However, the basin of attraction of initial guesses that converge to a given cycle is nevertheless very large, with the spread in acceptable initial guesses for figure 34.1 of order 1, in contrast to the exponential precision required of initial guesses by the Newton-Raphson method.



example 34.1
p. 707

exercise 34.3

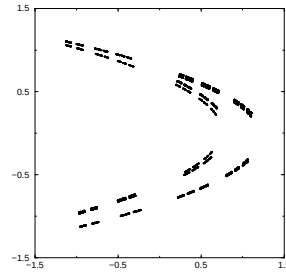


Figure 34.2: The repeller for the Hénon map at $a = 1.8$, $b = 0.3$.

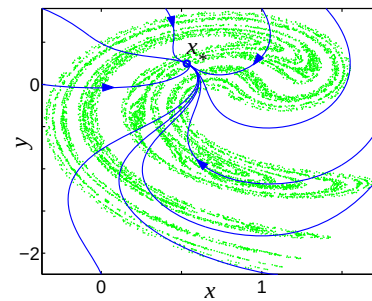


Figure 34.3: Typical trajectories of the vector field (34.9) for the stabilization of a hyperbolic fixed point of the Ikeda map (34.16) located at $(x, y) \approx (0.53275, 0.24689)$. The circle indicates the position of the fixed point. Note that the basin of attraction of this fixed point is large, larger than the entire Ikeda attractor.

The idea of the relaxation algorithm illustrated by the above Hénon map example is that instead of searching for an unstable periodic orbit of a map, one searches for a stable attractor of a vector field. More generally, consider a d -dimensional map $x' = f(x)$ with a hyperbolic fixed point x_* . Any fixed point x_* is by construction an equilibrium point of the fictitious time flow

$$\frac{dx}{d\tau} = f(x) - x. \quad (34.9)$$

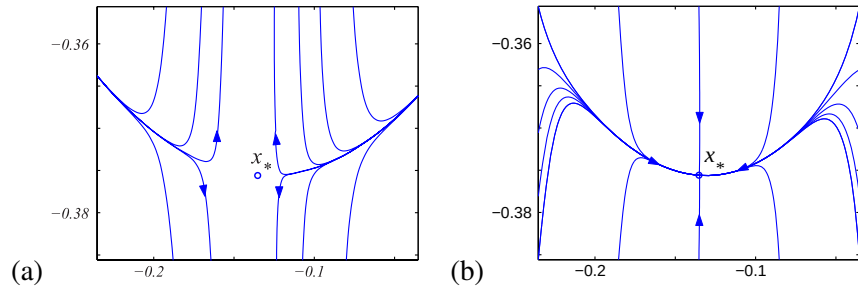
If all eigenvalues of the Jacobian matrix $J(x_*) = Df(x_*)$ have real parts smaller than unity, then x_* is a stable equilibrium point of the flow, see figure 34.3.

If some of the eigenvalues have real parts larger than unity, then one needs to modify the vector field so that the corresponding directions of the flow are turned into stable directions in a neighborhood of the fixed point. In the spirit of (34.8), modify the flow by

$$\frac{dx}{d\tau} = \mathbf{C}(f(x) - x), \quad (34.10)$$

where $\mathbf{C}$ is a $[d \times d]$ invertible matrix. The aim is to turn x_* into a stable equilibrium point of the flow by an appropriate choice of $\mathbf{C}$. It can be shown that a set of permutation / reflection matrices with one and only one non-vanishing entry ± 1 per row or column (for d -dimensional systems, there are $d!2^d$ such matrices) suffices to stabilize any fixed point. In practice, one chooses a particular matrix $\mathbf{C}$, and the flow is integrated. For each choice of $\mathbf{C}$, one or more hyperbolic fixed points of the map may turn into stable equilibria of the flow, see figure 34.4.

Figure 34.4: Typical trajectories of the vector field (34.10) for a hyperbolic fixed point $(x, y) \approx (-0.13529, -0.37559)$ of f^3 , where f is the Ikeda map (34.16). The circle indicates the position of the fixed point. For the vector field corresponding to (a) $\mathbf{C} = \mathbf{I}$, x_* is a hyperbolic equilibrium point of the flow, while for (b) $\mathbf{C} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, x_* is an attracting equilibrium point.



 example 34.2
p. 707

The generalization from searches for fixed points to searches for cycles is straightforward. In order to determine a prime cycle $x = (x_1, x_2, \dots, x_n)$ of a d -dimensional map $x' = f(x)$, we modify the multipoint shooting method of sect. 16.2, and consider the nd -dimensional vector field

$$\frac{dx}{d\tau} = \mathbf{C}(f(x) - x), \quad (34.11)$$

where $f(x) = (f(x_n), f(x_1), f(x_2), \dots, f(x_{n-1}))$, and $\mathbf{C}$ is an invertible $[nd \times nd]$ matrix. For the Hénon map, it is sufficient to consider a set of 2^n diagonal matrices with eigenvalues ± 1 . Risking a bit of confusion, we denote by x , $f(x)$ both the d -dimensional vectors in (34.10), and nd -dimensional vectors in (34.11), as the structure of the equations is the same.

34.2 Discrete iteration relaxation method

(C. Chandre, F.K. Diakonov and P. Schmelcher)

The problem with the Newton-Raphson iteration (34.2) is that it requires very precise initial guesses. For example, the n th iterate of a unimodal map has as many as 2^n periodic points crammed into the unit interval, so determination of all cycles of period n requires that the initial guess for each one of them has to be accurate to roughly 2^{-n} . This is not much of a problem for 1-dimensional maps, but making a good initial guess for where a cycle might lie in a d -dimensional state space can be a challenge.

Emboldened by the success of the cyclist relaxation trick (34.8) of manually turning instability into stability by a sign change, we now (i) abandon the Newton-Raphson method altogether, (ii) abandon the continuous fictitious time flow (34.9) with its time-consuming integration, replacing it by a map g with a larger basin of attraction (not restricted to a linear neighborhood of the fixed point). The idea is to construct a very simple map g , a linear transformation of the original f , for which the fixed point is stable. We replace the Jacobian matrix prefactor in (34.2) (whose inversion can be time-consuming) by a constant matrix prefactor

$$x' = g(x) = x + \Delta\tau \mathbf{C}(f(x) - x), \quad (34.12)$$

where $\Delta\tau$ is a positive real number, and $\mathbf{C}$ is a $[d \times d]$ permutation and reflection matrix with one and only one non-vanishing entry ± 1 per row or column. A fixed point of f is also a fixed point of g . Since $\mathbf{C}$ is invertible, the inverse is also true.

This construction is motivated by the observation that for small $\Delta\tau \rightarrow d\tau$ the map (34.12) is the Euler method for integrating the modified flow (34.10), with the integration step $\Delta\tau$.

The argument why a suitable choice of matrix $\mathbf{C}$ can lead to the stabilization of an unstable periodic orbit is similar to the one used to motivate the construction of the modified vector field in sect. 34.1. Indeed, the flow (34.8) is the simplest example of this method, with the infinitesimal fictitious time increment $\Delta\tau \rightarrow d\tau$, the infinitesimal coordinate correction $(x - x') \rightarrow dx_i$, and the $[n \times n]$ diagonal matrix $\mathbf{C} \rightarrow \sigma_i = \pm 1$.

For a given fixed point of $f(x)$ we again chose a $\mathbf{C}$ such that the flow in the expanding directions of $M(x_*)$ is turned into a contracting flow. The aim is to stabilize x_* by a suitable choice of $\mathbf{C}$. In the case where the map has multiple fixed points, the set of fixed points is obtained by changing the matrix $\mathbf{C}$ (in general different for each unstable fixed point) and varying initial conditions for the map g . For example, for 2-dimensional dissipative maps it can be shown that the 3

remark 34.2

$$\mathbf{C} \in \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right\}$$

suffice to stabilize all kinds of possible hyperbolic fixed points.

If $\Delta\tau$ is chosen sufficiently small, the magnitude of the eigenvalues of the fixed point x_* in the transformed system are smaller than one, and one has a stable fixed point. However, $\Delta\tau$ should not be chosen too small: Since the convergence is geometrical with a ratio $1 - \alpha\Delta\tau$ (where the value of constant α depends on the stability of the fixed point in the original system), small $\Delta\tau$ can slow down the speed of convergence. The critical value of $\Delta\tau$, which just suffices to make the fixed point stable, can be read off from the quadratic equations relating the stability coefficients of the original system and those of the transformed system. In practice, one can find the optimal $\Delta\tau$ by iterating the dynamical system stabilized with a given $\mathbf{C}$ and $\Delta\tau$. In general, all starting points converge on the attractor provided $\Delta\tau$ is small enough. If this is not the case, the trajectory either diverges (if $\Delta\tau$ is far too large) or it oscillates in a small section of the state space (if $\Delta\tau$ is close to its stabilizing value).

The search for the fixed points is now straightforward: A starting point chosen in the global neighborhood of the fixed point iterated with the transformed dynamical system g converges to the fixed point due to its stability. Numerical investigations show that the domain of attraction of a stabilized fixed point is a rather extended connected area, by no means confined to a linear neighborhood. At times the basin of attraction encompasses the complete state space of the attractor, so one can be sure to be within the attracting basin of a fixed point regardless of where on the attractor one picks the initial condition.

The step size $|g(x) - x|$ decreases exponentially when the trajectory approaches the fixed point. To get the coordinates of the fixed points with a high precision, one therefore needs a large number of iterations for the trajectory which is already in the linear neighborhood of the fixed point. To speed up the convergence of the final part of the approach to a fixed point we recommend a combination of the above approach with the Newton-Raphson method (34.2).

The fixed points of the n th iterate f^n are periodic points of a cycle of period n . If we consider the map

$$x' = g(x) = x + \Delta\tau \mathbf{C}(f^n(x) - x), \quad (34.13)$$

the iterates of g converge to a fixed point provided that $\Delta\tau$ is sufficiently small and $\mathbf{C}$ is a $[d \times d]$ constant matrix chosen such that it stabilizes the flow. As n grows, $\Delta\tau$ has to be chosen smaller and smaller. In the case of the Ikeda map example 34.2 the method works well for $n \leq 20$. As in (34.11), the multipoint shooting method is the method of preference for determining longer cycles. Consider $x = (x_1, x_2, \dots, x_n)$ and the nd -dimensional map

$$x' = f(x) = (f(x_n), f(x_1), \dots, f(x_{n-1})).$$

Determining cycles with period n for the d -dimensional f is equivalent to determining fixed points of the multipoint dn -dimensional f . The idea is to construct a matrix $\mathbf{C}$ such that the fixed point of f becomes stable for the map:

$$x' = x + \Delta\tau \mathbf{C}(f(x) - x),$$

where $\mathbf{C}$ is now a $[nd \times nd]$ permutation/reflection matrix with only one non-zero matrix element ± 1 per row or column. For any given matrix $\mathbf{C}$, a certain fraction of the cycles becomes stable and can be found by iterating the transformed map which is now a nd dimensional map.

From a practical point of view, the main advantage of this method compared to the Newton-Raphson method is twofold: (i) the Jacobian matrix of the flow need not be computed, so there is no large matrix to invert, simplifying considerably the implementation, and (ii) empirical basins of attractions for individual $\mathbf{C}$ are much larger than for the Newton-Raphson method. The price is a reduction in the speed of convergence.

34.3 Least action method

(P. Dahlqvist)

The methods of sects. 34.1 and 34.2 are somewhat *ad hoc*, as for general flows and iterated maps there is no fundamental principle to guide us in choosing the cost function, such as (34.3), to vary.

For Hamiltonian dynamics, we are on much firmer ground; Maupertuis least action principle. You yawn your way through it in every mechanics course—but as we shall now see, it is a very hands-on numerical method for finding cycles.

Table 34.1: All prime cycles up to 6 bounces for the 3-disk fundamental domain, center-to-center separation $R = 6$, disk radius $a = 1$. The columns list the cycle itinerary, its expanding eigenvalue Λ_p , and the length of the orbit (if the velocity=1 this is the same as its period or the action). Note that the two 6 cycles 001011 and 001101 are degenerate due to the time reversal symmetry, but are not related by any discrete spatial symmetry. (Computed by P.E. Rosenqvist.)

p	Λ_p	T_p
0	9.898979485566	4.000000000000
1	$-1.177145519638 \times 10^1$	4.267949192431
01	$-1.240948019921 \times 10^2$	8.316529485168
001	$-1.240542557041 \times 10^3$	12.321746616182
011	$1.449545074956 \times 10^3$	12.580807741032
0001	$-1.229570686196 \times 10^4$	16.322276474382
0011	$1.445997591902 \times 10^4$	16.585242906081
0111	$-1.707901900894 \times 10^4$	16.849071859224
00001	$-1.217338387051 \times 10^5$	20.322330025739
00011	$1.432820951544 \times 10^5$	20.585689671758
00101	$1.539257907420 \times 10^5$	20.638238386018
00111	$-1.704107155425 \times 10^5$	20.853571517227
01011	$-1.799019479426 \times 10^5$	20.897369388186
01111	$2.010247347433 \times 10^5$	21.116994322373
000001	$-1.205062923819 \times 10^6$	24.322335435738
000011	$1.418521622814 \times 10^6$	24.585734788507
000101	$1.525597448217 \times 10^6$	24.638760250323
000111	$-1.688624934257 \times 10^6$	24.854025100071
001011	$-1.796354939785 \times 10^6$	24.902167001066
001101	$-1.796354939785 \times 10^6$	24.902167001066
001111	$2.005733106218 \times 10^6$	25.121488488111
010111	$2.119615015369 \times 10^6$	25.165628236279
011111	$-2.366378254801 \times 10^6$	25.384945785676

Indeed, the simplest and numerically most robust method for determining cycles of planar billiards is given by the principle of least action, or equivalently, by extremizing the length of an approximate orbit that visits a given sequence of disks. In contrast to the multipoint shooting method of sect. 16.2 which requires variation of $2n$ phase-space points, extremization of a cycle length requires variation of only n bounce positions s_i .

The problem is to find the extremum values of cycle length $L(s)$ where $s = (s_1, \dots, s_n)$, that is find the roots of $\partial_i L(s) = 0$. Expand to first order

$$\partial_i L(s_0 + \delta s) = \partial_i L(s_0) + \sum_j \partial_i \partial_j L(s_0) \delta s_j + \dots$$

and use $M_{ij}(s_0) = \partial_i \partial_j L(s_0)$ in the n -dimensional Newton-Raphson iteration scheme of sect. 7.1.2

exercise 34.1

$$s_i \mapsto s_i - \sum_j \left(\frac{1}{M(s)} \right)_{ij} \partial_j L(s) \quad (34.14)$$

The extremization is achieved by recursive implementation of the above algorithm, with proviso that if the dynamics is pruned, one also has to check that the final extremal length orbit does not penetrate a billiard wall.

exercise 34.2
exercise 16.11

As an example, the short periods and stabilities of 3-disk cycles computed this way are listed table 34.1.

Résumé

Unlike the Newton-Raphson method, variational methods are very robust. As each step around a cycle is short, they do not suffer from exponential instabilities, and with rather coarse initial guesses one can determine cycles of arbitrary period.

Commentary

Remark 34.1. Relaxation method. The relaxation (or gradient) algorithm is one of the methods for solving extremal problems [24]. The method described above was introduced by Biham and Wenzel [2], who have also generalized it (in the case of the Hénon map) to determination of *all* 2^n cycles of period n , real or complex [3]. The applicability and reliability of the method is discussed in detail by Grassberger, Kantz and Moening [12], who give examples of the ways in which the method fails: (a) it might reach a limit cycle rather than an equilibrium saddle point (that can be remedied by the complex Biham-Wenzel algorithm [3]) (b) different symbol sequences can converge to the same cycle (i.e., more refined initial conditions might be needed). Furthermore, Hansen (ref. [13] and chapter 4. of ref. [14]) has pointed out that the method cannot find certain cycles for specific values of the Hénon map parameters. In practice, the relaxation method for determining periodic orbits of maps appears to be effective almost always, but not always. It is much slower than the multipoint shooting method of sect. 16.2, but also much quicker to program, as it does not require evaluation of stability matrices and their inversion. If the complete set of cycles is required, the method has to be supplemented by other methods.

Remark 34.2. Hybrid Newton-Raphson/relaxation methods. The method discussed in sect. 34.2 was introduced by Schmelcher *et al* [9, 21]. The method was extended to flows by means of the Poincaré section technique in ref. [20]. It is also possible to combine the Newton-Raphson method and (34.12) in the construction of a transformed map [6]. In this approach, each step of the iteration scheme is a linear superposition of a step of the stability transformed system and a step of the Newton-Raphson algorithm. Far from the linear neighborhood the weight is dominantly on the globally acting stability transformation algorithm. Close to the fixed point, the steps of the iteration are dominated by the Newton-Raphson procedure.

Remark 34.3. Relation to the Smale horseshoe symbolic dynamics. For a complete horseshoe Hénon repeller (a sufficiently large), such as the one given in figure 34.2, the signs $\sigma_i \in \{1, -1\}$ are in a 1-to-1 correspondence with the Smale horseshoe symbolic dynamics $s_i \in \{0, 1\}$:

$$s_i = \begin{cases} 0 & \text{if } \sigma_i = -1, \quad x_i < 0 \\ 1 & \text{if } \sigma_i = +1, \quad x_i > 0 \end{cases} \quad (34.15)$$

For arbitrary parameter values with a finite subshift symbolic dynamics or with arbitrarily complicated pruning, the relation of sign sequences $\{\sigma_1, \sigma_2, \dots, \sigma_n\}$ to the itineraries $\{s_1, s_2, \dots, s_n\}$ can be much subtler; this is discussed in ref. [12].

Remark 34.4. Ikeda map. Ikeda map (34.16) was introduced in ref. [15] is a model which exhibits complex dynamics observed in nonlinear optical ring cavities.

Remark 34.5. Relaxation for continuous time flows. For a d -dimensional flow $\dot{x} = v(x)$, the method described above can be extended by considering a Poincaré section. The Poincaré section yields a map f with dimension $d-1$, and the above discrete iterative maps procedures can be carried out. A method that keeps the trial orbit continuous throughout the calculation is the Newton descent, a variational method for finding periodic orbits of continuous time flows, is described in refs. [4, 17].

Remark 34.6. Stability ordering. The parameter $\Delta\tau$ in (34.12) is a key quantity here. It is related to the stability of the desired cycle in the transformed system: The more unstable a fixed point is, the smaller $\Delta\tau$ has to be to stabilize it. With increasing cycle periods, the unstable eigenvalue of the Jacobian matrix increases and therefore $\Delta\tau$ has to be reduced to achieve stabilization of all fixed points. In many cases the least unstable cycles of a given period n are of physically most important [9]. In this context $\Delta\tau$ operates as a stability filter. It allows the selective stabilization of only those cycles which posses Lyapunov exponents smaller than a cut-off value. If one starts the search for cycles within a given period n with a value $\Delta\tau \approx O(10^{-1})$, and gradually lowers $\Delta\tau$ one obtains the sequence of all unstable orbits of order n sorted with increasing values of their Lyapunov exponents. For the specific choice of $\mathbf{C}$ the relation between $\Delta\tau$ and the stability coefficients of the fixed points of the original system is strictly monotonous. Transformed dynamical systems with other $\mathbf{C}$'s do not obey such a strict behavior but show a rough ordering of the sequence of Floquet multipliers of the fixed points stabilized in the course of decreasing values for $\Delta\tau$. As explained in sect. 23.7, stability ordered cycles are needed to order cycle expansions of dynamical quantities of chaotic systems for which a symbolic dynamics is not known. For such systems, an ordering of cycles with respect to their stability has been proposed [5, 7, 8], and shown to yield good results in practical applications.

section 23.7

Remark 34.7. Action extremization method. The action extremization (sect. 34.3) as a numerical method for finding cycles has been introduced independently by many people. We have learned it from G. Russberg, and from M. Sieber’s and F. Steiner’s hyperbola billiard computations [22, 23]. The convergence rate is really impressive, for the Sinai billiard some 5000 cycles are computed within CPU seconds with rather bad initial guesses.

Variational methods are the key ingredient of the Aubry-Mather theory of area-preserving twist maps (known in the condensed matter literature as the Frenkel-Kontorova models of 1-dimensional crystals), discrete-time Hamiltonian dynamical systems particularly suited to explorations of the K.A.M. theorem. Proofs of the Aubry-Mather theorem [19] on existence of quasi-periodic solutions are variational. It was quickly realized that the variational methods can also yield reliable, high precision computations of long periodic orbits of twist map models in 2 or more dimensions, needed for K.A.M. renormalization studies [16].

A fictitious time gradient flow similar to the one discussed here in sect. 34.1 was introduced by Anagnost [1] for twist maps, and used by Gole [11] in his proof of the Aubry-Mather theorem. Mathematical bounds on the regions of stability of K.A.M. tori are notoriously restrictive compared to the numerical indications, and de la Llave, Falcolini and Tompaids [10, 25] have found the gradient flow formulation advantageous both in studies of the analyticity domains of the K.A.M. stability, as well as proving the Aubry-Mather theorem for extended systems (for a pedagogical introduction, see the lattice dynamics section of ref. [18]).

All of the twist-maps work is based on extremizing the discrete dynamics version of the action S (in this context sometimes called a “generating function”). However, in their investigations in the complex plane, Falcolini and de la Llave [10] do find it useful to minimize instead $S\bar{S}$, analogous to our cost function (34.3).

References

- [1] S. B. Angenent, “The periodic orbits of an area preserving twist-map”, *Commun. Math. Phys.* **115**, 353–374 (1988).
- [2] O. Biham and W. Wenzel, “Characterization of unstable periodic orbits in chaotic attractors and repellers”, *Phys. Rev. Lett.* **63**, 819–822 (1989).
- [3] O. Biham and W. Wenzel, “Unstable periodic orbits and the symbolic dynamics of the complex Hénon map”, *Phys. Rev. E* **42**, 4639 (1990).
- [4] P. Cvitanović and Y. Lan, Turbulent fields and their recurrences, in *Correlations and Fluctuations in QCD : Proceedings of 10. International Workshop on Multiparticle Production*, edited by N. Antoniou (2003), pp. 313–325.
- [5] P. Dahlqvist and G. Russberg, “Periodic orbit quantization of bound chaotic systems”, *J. Phys. A* **24**, 4763–4778 (1991).
- [6] R. L. Davidchack and Y.-C. Lai, “Efficient algorithm for detecting unstable periodic orbits in chaotic systems”, *Phys. Rev. E* **60**, 6172 (1999).
- [7] C. P. Dettmann and P. Cvitanović, “Cycle expansions for intermittent diffusion”, *Phys. Rev. E* **56**, 6687–6692 (1997).

- [8] C. P. Dettmann and G. P. Morriss, “Stability ordering; strong field Lorentz gas”, *Phys. Rev. Lett.* **78**, 4201–4204 (1997).
- [9] F. K. Diakonov, P. Schmelcher, and O. Biham, “Systematic computation of the least unstable periodic orbits in chaotic attractors”, *Phys. Rev. Lett.* **81**, 4349 (1998).
- [10] C. Falcolini and R. de la Llave, “Numerical calculation of domains of analyticity for perturbation theories in the presence of small divisors”, *J. Stat. Phys.* **67**, 645–666 (1992).
- [11] C. Golé, “A new proof of the Aubry-Mather’s theorem”, *Math. Z.* **210**, 441–448 (1992).
- [12] P. Grassberger, H. Kantz, and U. Moenig, “On the symbolic dynamics of Hénon map”, *J. Phys. A* **22**, 5217–5230 (1989).
- [13] K. T. Hansen, “Remarks on the symbolic dynamics for the Hénon map”, *Phys. Lett. A* **165**, 100–104 (1992).
- [14] K. T. Hansen, *Symbolic Dynamics in Chaotic systems*, PhD thesis (Univ. of Oslo, Oslo, Norway, 1993).
- [15] K. Ikeda, “Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system”, *Opt. Commun.* **30**, 257–261 (1979).
- [16] H. Kook and J. D. Meiss, “Periodic orbits for reversible, symplectic mappings”, *Physica D* **35**, 65–86 (1989).
- [17] Y. Lan and P. Cvitanović, “Variational method for finding periodic orbits in a general flow”, *Phys. Rev. E* **69**, 016217 (2004).
- [18] R. de la Llave, *Variational methods for quasiperiodic solutions of partial differential equations*, in *Hamiltonian Systems and Celestial Mechanics (HAMSYS-98)*, edited by J. Delgado, E. A. Lacomba, E. Pérez-Chavela, and J. Llibre (2000).
- [19] J. N. Mather, “Variational construction of orbits of twist diffeomorphisms”, *J. Am. Math. Soc.* **4**, 207–263 (1991).
- [20] D. Pingel, P. Schmelcher, F. K. Diakonov, and O. Biham, “Theory and applications of the systematic detection of unstable periodic orbits in dynamical systems”, *Phys. Rev. E* **62**, 2119 (2000).
- [21] P. Schmelcher and F. K. Diakonov, “General approach to the localization of unstable periodic orbits in chaotic dynamical systems”, *Phys. Rev. E* **57**, 2739 (1998).
- [22] M. Sieber, *The Hyperbola Billiard: A Model for the Semiclassical Quantization of Chaotic Systems*, DESY report 91-030, PhD thesis (Hamburg Univ., 1991).
- [23] M. Sieber and F. Steiner, “Quantum chaos in the hyperbola billiard”, *Phys. Lett. A* **148**, 415–420 (1990).
- [24] F. Stummel and K. Hainer, *Praktische Mathematik* (Vieweg+Teubner, 1982).
- [25] S. Tompaids, “Numerical study of invariant sets of a quasi-periodic perturbation of a symplectic map”, *Experiment. Math.* **5**, 211–230 (1996).

34.4 Examples

Example 34.1. Hénon map cycles. Our aim in this calculation is to find all periodic orbits of period n for the Hénon map (34.4), in principle at most 2^n orbits. We start by choosing an initial guess trajectory $(x_1, x_2, \dots, x_n)$ and impose the periodic boundary condition $x_{n+1} = x_1$. The simplest and a rather crude choice of the initial condition in the Hénon map example is $x_i = 0$ for all i . In order to find a given orbit one sets $\sigma_i = -1$ for all iterates i which are local minima of $V_i(x)$, and $\sigma_i = 1$ for iterates which are local maxima. In practice one runs through a complete list of prime cycles, such as the table 18.1. The real issue for all searches for periodic orbits, this one included, is how large is the basin of attraction of the desired periodic orbit? There is no easy answer to this question, but empirically it turns out that for the Hénon map such initial guess almost always converges to the desired trajectory as long as the initial $|x|$ is not too large compared to $1/\sqrt{a}$. Figure 34.1 gives some indication of a typical basin of attraction of the method (see also figure 34.3).

The calculation is carried out by solving the set of n ordinary differential equations (34.8) using a simple Runge-Kutta method with a relatively large step size ($h = 0.1$) until $|v|$ becomes smaller than a given value ε (in a typical calculation $\varepsilon \sim 10^{-7}$). Empirically, in the case that an orbit corresponding to the desired itinerary does not exist, the initial guess escapes to infinity since the “potential” $V_i(x)$ grows without bound.

exercise 34.3

Applied to the Hénon map at the Hénon’s parameters choice $a = 1.4$, $b = 0.3$, the method has yielded all periodic orbits to periods as long as $n = 28$, as well as selected orbits up to period $n = 1000$. All prime cycles up to period 10 for the Hénon map, $a = 1.4$ and $b = 0.3$, are listed in table 34.2. The number of unstable periodic orbits for periods $n \leq 28$ is given in table 34.3. Comparing this with the list of all possible 2-symbol alphabet prime cycles, table 18.1, we see that the pruning is quite extensive, with the number of periodic points of period n growing as $e^{0.4645 \cdot n} = (1.592)^n$ rather than as 2^n .

As another example we plot all unstable periodic points up to period $n = 14$ for $a = 1.8$, $b = 0.3$ in figure 34.2. Comparing this repelling set with the strange attractor for the Hénon’s parameters figure 3.7, we note the existence of gaps in the set, cut out by the preimages of the escaping regions.

remark 34.1

In practice, the relaxation flow (34.8) finds (almost) all periodic orbits which exist and indicates which ones do not. For the Hénon map the method enables us to calculate almost all unstable cycles of essentially any desired period and accuracy.

return to p. 697

Example 34.2. Ikeda map: We illustrate the method of (34.10) with the determination of the periodic orbits of the Ikeda map:

$$\begin{aligned} x' &= 1 + a(x \cos w - y \sin w) \\ y' &= a(x \sin w + y \cos w) \\ \text{where } w &= b - \frac{c}{1 + x^2 + y^2}, \end{aligned} \quad (34.16)$$

with $a = 0.9$, $b = 0.4$, $c = 6$. The fixed point x_* is located at $(x, y) \approx (0.53275, 0.24689)$, with eigenvalues of the Jacobian matrix $(\Lambda_1, \Lambda_2) \approx (-2.3897, -0.3389)$, so the flow is already stabilized with $C = 1$. Figure 34.3 depicts the flow of the vector field around the fixed point x_* .

Table 34.2: All prime cycles up to period 10 for the Hénon map, $a = 1.4$ and $b = 0.3$. The columns list the period n_p , the itinerary (defined in remark 34.3), a periodic point (y_p, x_p) , and the cycle Lyapunov exponent $\lambda_p = \ln |\Lambda_p|/n_p$. While most of the cycles have $\lambda_p \approx 0.5$, several significantly do not. The $\bar{0}$ periodic point is very unstable, isolated and transient fixed point, with no other cycles returning close to it. At period 13 one finds a pair of cycles with exceptionally low Lyapunov exponents. The cycles are close for most of the trajectory, differing only in the one symbol corresponding to two periodic points straddle the (partition) fold of the attractor. As the system is not hyperbolic, there is no known lower bound on cycle Lyapunov exponents, and the Hénon's strange "attractor" might some day turn out to be nothing but a transient on the way to a periodic attractor of some long period.

n	p	(y_p, x_p)	λ_p
1	0	(-1.13135447, -1.13135447)	1.18167262
	1	(0.63135447, 0.63135447)	0.65427061
2	01	(0.97580005, -0.47580005)	0.55098676
4	0111	(-0.70676677, 0.63819399)	0.53908457
6	010111	(-0.41515894, 1.07011813)	0.55610982
	011111	(-0.80421990, 0.44190995)	0.55245341
7	0011101	(-1.04667757, -0.17877958)	0.40998559
	0011111	(-1.08728604, -0.28539206)	0.46539757
	0101111	(-0.34267842, 1.14123046)	0.41283650
	0111111	(-0.88050537, 0.26827759)	0.51090634
8	00011101	(-1.25487963, -0.82745422)	0.43876727
	00011111	(-1.25872451, -0.83714168)	0.43942101
	00111101	(-1.14931330, -0.48368863)	0.47834615
	00111111	(-1.14078564, -0.44837319)	0.49353764
	01010111	(-0.52309999, 0.93830866)	0.54805453
	01011111	(-0.38817041, 1.09945313)	0.55972495
	01111111	(-0.83680827, 0.36978609)	0.56236493
9	000111101	(-1.27793296, -0.90626780)	0.38732115
	000111111	(-1.27771933, -0.90378859)	0.39621864
	001111101	(-1.10392601, -0.34524675)	0.51112950
	001111111	(-1.11352304, -0.36427104)	0.51757012
	010111111	(-0.36894919, 1.11803210)	0.54264571
	011111111	(-0.85789748, 0.32147653)	0.56016658
10	0001111101	(-1.26640530, -0.86684837)	0.47738235
	0001111111	(-1.26782752, -0.86878943)	0.47745508
	0011111101	(-1.12796804, -0.41787432)	0.52544529
	0011111111	(-1.12760083, -0.40742737)	0.53063973
	0101010111	(-0.48815908, 0.98458725)	0.54989554
	0101011111	(-0.53496022, 0.92336925)	0.54960607
	0101110111	(-0.42726915, 1.05695851)	0.54836764
	0101111111	(-0.37947780, 1.10801373)	0.56915950
	0111011111	(-0.69555680, 0.66088560)	0.54443884
	0111111111	(-0.84660200, 0.34750875)	0.57591048
13	1110011101000	(-1.2085766485, -0.6729999948)	0.19882434
	1110011101001	(-1.0598110494, -0.2056310390)	0.21072511

Table 34.3: The number of unstable periodic orbits of the Hénon map for $a = 1.4$, $b = 0.3$, of all periods $n \leq 28$. M_n is the number of prime cycles of period n , and N_n is the total number of periodic points of period n (including repeats of shorter prime cycles).

n	M_n	N_n	n	M_n	N_n	n	M_n	N_n
11	14	156	17	166	2824	23	1930	44392
12	19	248	18	233	4264	24	2902	69952
13	32	418	19	364	6918	25	4498	112452
14	44	648	20	535	10808	26	6806	177376
15	72	1082	21	834	17544	27	10518	284042
16	102	1696	22	1225	27108	28	16031	449520

In order to determine x_* , one needs to integrate the vector field (34.9) forward in time (the convergence is exponential in time), using a fourth order Runge-Kutta or any other integration routine.

In contrast, determination of the 3-cycles of the Ikeda map requires nontrivial $\mathbf{C}$ matrices, different from the identity. Consider for example the hyperbolic fixed point $(x, y) \approx (-0.13529, -0.37559)$ of the third iterate f^3 of the Ikeda map. The flow of the vector field for $\mathbf{C} = \mathbf{I}$, Figure 34.4 (a), indicates a hyperbolic equilibrium point, while for $\mathbf{C} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ the flow of the vector field, figure 34.4 (b) indicates that x_* is an attracting equilibrium point, reached at exponential speed by integration forward in time.

[return to p. 699](#)

Exercises

- 34.1. **Evaluation of billiard cycles by minimization***. Given a symbol sequence, you can construct a guess trajectory by taking a point on the boundary of each disk in the sequence, and connecting them by straight lines. If this were a rubber band wrapped through 3 rings, it would shrink into the physical trajectory, which minimizes the action (in this case, the length) of the trajectory.
- Write a program to find the periodic orbits for your billiard simulator. Use the least action principle to extremize the length of the periodic orbit, and reproduce the periods and stabilities of 3-disk cycles, table 34.1. (One such method is given in sect. 34.3.) After that check the accuracy of the computed orbits by iterating them forward with your simulator. What is your error $|f^{T_p}(x) - x|$?
- 34.2. **Tracking cycles adiabatically***. Once a cycle has been found, orbits for different system parameters values may be obtained by varying slowly (adiabatically) the parameters, and using the old orbit points as starting guesses in the Newton method. Try this method out on the 3-disk system. It works well for $R : a$ sufficiently large. For smaller values, some orbits change rather quickly and require very small step sizes. In addition, for ratios below $R : a = 2.04821419 \dots$ families of cycles are pruned, i.e. some of the minimal length trajectories are blocked by intervening disks.
- 34.3. **Cycles of the Hénon map.** Apply the method of sect. 34.1 to the Hénon map at the Hénon's parameters choice $a = 1.4$, $b = 0.3$, and compute all prime cycles for at least $n \leq 6$. Estimate the topological entropy, either from the definition (18.1), or as the zero of a truncated topological zeta function (18.17). Do your cycles agree with the cycles listed in table 34.2?

Part V

Quantum chaos

YOU HAVE MASTERED part II of ChaosBook. You can play a game of pinball, and if you are a skilled neuroscientist, you now know how to poke rat brains. You have learned that information about chaotic dynamics can be obtained by calculating spectra of linear operators such as the evolution operator, and that these spectra can be expressed in terms of periodic orbits by means of cycle expansions.

But what happens if we scatter quantum mechanical waves rather than point-like pinballs? Is there a link between quantum-mechanical spectra and the dynamics of the underlying deterministic flow? The answer is yes, in a very pleasing way - essentially the same ζ functions and cycle expansions describe the deterministic chaotic dynamics, the stochastic dynamics, and the semiclassical quantum mechanics (chapter 35).

1. We start with a lightning review of quantum mechanics (chapter 36) and then discuss the first semiclassical (or WKB) approach to quantization (chapter 37).
2. Then the semiclassical evolution operator (chapter 38) leads to the semiclassical trace formulas and ζ functions quantization formulas (chapter 39).
3. Their simplest applications are through trace formulas for scattering (chapter 40) and multi-scattering (chapter 41).
4. Now that we have derived the semiclassical weight associated with every unstable periodic orbit, we are now able to put together all ingredients that make the game of pinball unpredictable, and compute a “chaotic” part of the helium spectrum to shocking accuracy (chapter 42).
5. A semiclassical theory in terms of deterministic dynamics alone cannot be exact. Waves interfere, diffract (chapter 43), and higher $\hbar$ corrections need to be incorporated into the periodic orbit theory.

This part of ChaosBook is a collaborative effort of Predrag Cvitanović, Roberto Artuso, Per Dahlqvist, Ronnie Mainieri, Gregor Tanner, Gábor Vattay, Niall Whelan, and Andreas Wirzba.

Chapter 35

Prologue

Anyone who uses words “quantum” and “chaos” in the same sentence should be hung by his thumbs on a tree in the park behind the Niels Bohr Institute.

—Joseph Ford

(G. Vattay, G. Tanner and P. Cvitanović)

SO FAR we have learned that information about chaotic dynamics can be obtained by calculating spectra of linear operators such as the evolution operator of sect. 20.2 or the associated partial differential equations such as the Liouville equation (19.34). The spectra of these operators can be expressed in terms of periodic orbits of the deterministic dynamics by means of periodic orbit expansions.

But what happens quantum mechanically, i.e., if we scatter waves rather than point-like pinballs? Can we turn the problem round and study linear PDE's in terms of the underlying deterministic dynamics? And, is there a link between structures in the spectrum or the eigenfunctions of a PDE and the dynamical properties of the underlying deterministic flow? The answer is yes, but . . . things are becoming somewhat more complicated when studying 2nd or higher order linear PDE's. We can find deterministic dynamics associated with a linear PDE, just take geometric optics as a familiar example. Propagation of light follows a second order wave equation but may in certain limits be well described in terms of geometric rays. A theory in terms of properties of the deterministic dynamics alone, referred to here as the *semiclassical theory*, will not be exact, in contrast to the deterministic periodic orbit formulas obtained so far. Waves exhibit new phenomena, such as interference, diffraction, and higher $\hbar$ corrections which will only be partially incorporated into the periodic orbit theory.

chapter 43

35.1 Quantum pinball

In what follows, we will restrict the discussion to the non-relativistic Schrödinger equation. The approach will be very much in the spirit of the early days of quantum mechanics, before its wave character had been fully uncovered by Schrödinger in the mid 1920's. Indeed, if physicists of the period were as familiar with deterministic chaos as we are today, this theory could have been developed 80 years ago. It was the discrete nature of the hydrogen spectrum which inspired the Bohr-de Broglie picture of the old quantum theory: one places a wave instead of a particle on a Keplerian orbit around the hydrogen nucleus. The quantization condition is that only those orbits contribute for which this wave is stationary; from this followed the Balmer spectrum and the Bohr-Sommerfeld quantization which eventually led to the more sophisticated theory of Heisenberg, Schrödinger and others. Today we are very aware of the fact that elliptic orbits are an idiosyncrasy of the Kepler problem, and that chaos is the rule; so can the Bohr quantization be generalized to chaotic systems?

The question was answered affirmatively by M. Gutzwiller, as late as 1971; a chaotic system can indeed be quantized by placing a wave on each of the *infinity* of unstable periodic orbits. Due to the instability of the orbits the wave does not stay localized but leaks into neighborhoods of other periodic orbits. Contributions of different periodic orbits interfere and the quantization condition can no longer be attributed to a single periodic orbit. A coherent summation over the infinity of periodic orbit contributions gives the desired spectrum.

The pleasant surprise is that the zeros of the dynamical zeta function (1.10) derived in the context of deterministic chaotic dynamics,

chapter 22

$$1/\zeta(z) = \prod_p (1 - t_p),$$

also yield excellent estimates of *quantum* resonances, with the quantum amplitude associated with a given cycle approximated semiclassically by the weight

$$t_p = \frac{1}{|\Lambda_p|^{\frac{1}{2}}} e^{\frac{i}{\hbar} S_p - i\pi m_p/2}, \quad (35.1)$$

whose magnitude is the square root of the deterministic weight (22.9)

$$t_p = \frac{1}{|\Lambda_p|} e^{\beta A_p - s T_p},$$

and the phase is given by the Bohr-Sommerfeld action integral S_p , together with an additional topological phase m_p , the number of caustics along the periodic trajectory, points where the naive semiclassical approximation fails.

chapter 38

In this approach, the quantal spectra of deterministically chaotic dynamical systems are determined from the zeros of dynamical zeta functions, defined by cycle expansions of infinite products of form

chapter 23

$$1/\zeta = \prod_p (1 - t_p) = 1 - \sum_f t_f - \sum_k c_k \quad (35.2)$$

with weight t_p associated to every prime (non-repeating) periodic orbit (or *cycle*) p .

The key observation is that the chaotic dynamics is often organized around a few *fundamental* cycles. These short cycles capture the skeletal topology of the motion in the sense that any long orbit can approximately be pieced together from the fundamental cycles. In chapter 23 it was shown that for this reason the cycle expansion (35.2) is a highly convergent expansion dominated by short cycles grouped into *fundamental* contributions, with longer cycles contributing rapidly decreasing *curvature* corrections. Computations with dynamical zeta functions are rather straightforward; typically one determines periods and stabilities of a finite number of shortest periodic orbits, substitutes them into (35.2), and estimates the zeros of $1/\zeta$ from such polynomial approximations.

From the vantage point of the dynamical systems theory, the trace formulas (both the exact Selberg and the semiclassical Gutzwiller trace formula) fit into a general framework of replacing phase space averages by sums over periodic orbits. For deterministic hyperbolic systems this is possible since the invariant density can be represented by sum over all periodic orbits, with weights related to their instability. The semiclassical periodic orbit sums differ from the deterministic ones only in phase factors and stability weights; such differences may be traced back to the fact that in quantum mechanics the amplitudes rather than the probabilities are added.

chapter 39

The type of dynamics has a strong influence on the convergence of cycle expansions and the properties of quantal spectra; this necessitates development of different approaches for different types of dynamical behavior such as, on one hand, the strongly hyperbolic and, on the other hand, the intermittent dynamics of chapters 23 and 29. For generic nonhyperbolic systems (which we shall not discuss here), with mixed phase space and marginally stable orbits, periodic orbit summations are hard to control, and it is still not clear that the periodic orbit sums should necessarily be the computational method of choice.

chapter 23

chapter 29

Where is all this taking us? The goal of this part of the book is to demonstrate that the cycle expansions, developed so far in deterministic settings, are also a powerful tool for evaluation of *quantum* resonances of deterministically chaotic systems.

First, we shall warm up playing our game of pinball, this time in a quantum version. Were the game of pinball a closed system, quantum mechanically one would determine its stationary eigenfunctions and eigenenergies. For open systems one seeks instead complex resonances, where the imaginary part of the eigenenergy describes the rate at which the quantum wave function leaks out of the central scattering region. This will turn out to work well, except who truly wants to know accurately the resonances of a quantum pinball?

chapter 40

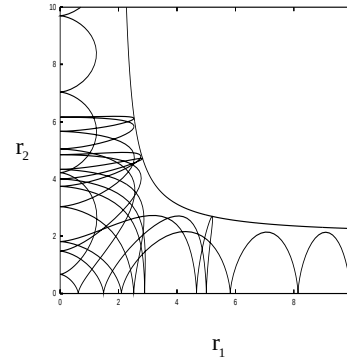


Figure 35.1: A typical collinear helium trajectory in the $r_1 - r_2$ plane; the trajectory enters along the r_1 axis and escapes to infinity along the r_2 axis.

35.2 Quantization of helium

Once we have derived the semiclassical (35.1) weight associated with the periodic orbit p , we will finally be in position to accomplish something altogether remarkable. We are now able to put together all ingredients that make the game of pinball unpredictable, and compute a “chaotic” part of the helium spectrum to shocking accuracy. From the deterministic dynamics point of view, helium is an example of Poincaré’s dreaded and intractable 3-body problem. Undaunted, we forge ahead and consider the *collinear* helium, with zero total angular momentum, and the two electrons on the opposite sides of the nucleus.



We set the electron mass to 1, the nucleus mass to ∞ , the helium nucleus charge to 2, the electron charges to -1. The Hamiltonian is

chapter 42

$$H = \frac{1}{2}p_1^2 + \frac{1}{2}p_2^2 - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_1 + r_2}. \quad (35.3)$$

Due to the energy conservation, only three of the phase space coordinates (r_1, r_2, p_1, p_2) are independent. The dynamics can be visualized as a motion in the (r_1, r_2) , $r_i \geq 0$ quadrant, figure 35.1, or, better still, by a well chosen 2-dimensional Poincaré section.

The motion in the (r_1, r_2) plane is topologically similar to the pinball motion in a 3-disk system, except that the motion is not free, but in the Coulomb potential. The deterministic collinear helium is also a repeller; almost all of the deterministic trajectories escape. Miraculously, the symbolic dynamics for the survivors turns out to be binary, just as in the 3-disk game of pinball, so we know what cycles need to be computed for the cycle expansion (1.11). A set of shortest cycles up to a given symbol string length then yields an estimate of the helium spectrum. This simple calculation yields surprisingly accurate eigenvalues; even though the cycle expansion was based on the *semiclassical approximation* (35.1) which is expected to be good only in the deterministic large energy limit, the eigenenergies are good to 1% all the way down to the ground state.

chapter 42

Before we can get to this point, we first have to recapitulate some basic notions

of quantum mechanics; after having defined the main quantum objects of interest, the quantum propagator and the Green's function, we will relate the quantum propagation to the deterministic flow of the underlying dynamical system. We will then proceed to construct semiclassical approximations to the quantum propagator and the Green's function. A rederivation of deterministic Hamiltonian dynamics starting from the Hamilton-Jacobi equation will be offered along the way. The derivation of the Gutzwiller trace formula and the semiclassical zeta function as a sum and as a product over periodic orbits will be given in chapter 39. In subsequent chapters we buttress our case by applying and extending the theory: a cycle expansion calculation of scattering resonances in a 3-disk billiard in chapter 40, the spectrum of helium in chapter 42, and the incorporation of diffraction effects in chapter 43.

Commentary

Remark 35.1. Guide to literature. A key prerequisite to developing any theory of “quantum chaos” is solid understanding of Hamiltonian mechanics. For that, Arnol'd's monograph [1] is the essential reference. Ozorio de Almeida's monograph [6] offers a compact introduction to the aspects of Hamiltonian dynamics required for the quantization of integrable and nearly integrable systems, with emphasis on periodic orbits, normal forms, catastrophe theory and torus quantization. The book by Brack and Bhaduri [2] is an excellent introduction to the semiclassical methods. Gutzwiller's monograph [3] is an advanced introduction focusing on chaotic dynamics both in deterministic Hamiltonian settings and in the semiclassical quantization. This book is worth browsing through for its many insights and erudite comments on quantum and celestial mechanics even if one is not working on problems of quantum chaos. More suitable as a graduate course texts are Stöckmann [9] (with emphasis on experiments) and Reichl's [7] expositions.

This book does not discuss the random matrix theory approach to chaos in quantal spectra; no randomness assumptions are made here, rather the goal is to milk the deterministic chaotic dynamics for its full worth. The book concentrates on the periodic orbit theory. For an introduction to “quantum chaos” that focuses on the random matrix theory the reader is referred to the excellent monograph by Haake [4], among others.

Remark 35.2. The dates. Schrödinger's first wave mechanics paper [8] (hydrogen spectrum) was submitted 27 January 1926. Submission date for Madelung's ‘quantum theory in hydrodynamical form’ paper [5] was 25 October 1926.

References

- [1] V. I. Arnol'd, *Mathematical Methods for Classical Mechanics* (Springer, New York, 1989).
- [2] M. Brack and R. Bhaduri, *Semiclassical Physics* (Westview Press, Boulder, CO, 2003).
- [3] M. C. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer, New York, 1990).

- [4] F. Haake, *Quantum Signatures of Chaos*, 3rd ed. (Springer, Berlin, 2010).
- [5] E. Madelung, “Quantentheorie in hydrodynamischer Form”, *Z. f. Physik* **40**, 322–326 (1927).
- [6] A. M. Ozorio de Almeida, *Hamiltonian Systems: Chaos and Quantization* (Cambridge Univ. Press, Cambridge UK, 1989).
- [7] L. Reichl, *The Transition to Chaos: Conservative Classical Systems and Quantum Manifestations*, 2nd ed. (Springer, New York, 2004).
- [8] E. Schrödinger, “Quantisierung als Eigenwertproblem”, *Ann. Phys.* **384**, 361–376 (1926).
- [9] H.-J. Stöckmann, *Quantum Chaos: An Introduction* (Cambridge Univ. Press, Cambridge UK, 2007).

Chapter 36

Quantum mechanics - the short short version

WE START WITH a review of standard quantum mechanical concepts prerequisite to the derivation of the semiclassical trace formula.

In coordinate representation, the time evolution of a quantum mechanical wave function is governed by the Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} \psi(q, t) = \hat{H} \left(q, \frac{\hbar}{i} \frac{\partial}{\partial q} \right) \psi(q, t), \quad (36.1)$$

where the Hamilton operator $\hat{H}(q, -i\hbar\partial_q)$ is obtained from the deterministic Hamiltonian by substituting $p \rightarrow -i\hbar\partial_q$. Most of the Hamiltonians we shall consider here are of the separable form

$$H(q, p) = T(p) + V(q), \quad T(p) = p^2/2m, \quad (36.2)$$

describing dynamics of a particle in a D -dimensional potential $V(q)$. For time-independent Hamiltonians we are interested in finding stationary solutions of the Schrödinger equation of the form

$$\psi_n(q, t) = e^{-iE_n t/\hbar} \phi_n(q), \quad (36.3)$$

where E_n are the eigenenergies of the time-independent Schrödinger equation

$$\hat{H}\phi(q) = E\phi(q). \quad (36.4)$$

For bound systems, the spectrum is discrete and the eigenfunctions form an orthonormal,

$$\int dq \phi_n(q) \phi_m^*(q) = \delta_{nm}, \quad (36.5)$$

and complete,

$$\sum_n \phi_n(q) \phi_n^*(q') = \delta(q - q'), \quad (36.6)$$

set of functions in a Hilbert space. Here and throughout the text,

$$\int dq = \int dq_1 dq_2 \dots dq_D. \quad (36.7)$$

For simplicity, we will assume that the system is bound, although most of the results will be applicable to open systems, where one has complex resonances instead of real energies, and the spectrum has continuous components. chapter 40

A given wave function can be expanded in the energy eigenbasis

$$\psi(q, t) = \sum_n c_n e^{-iE_n t/\hbar} \phi_n(q), \quad (36.8)$$

where the expansion coefficient c_n is given by the projection of the initial wave function $\psi(q, 0)$ onto the n th eigenstate

$$c_n = \int dq \phi_n^*(q) \psi(q, 0). \quad (36.9)$$

By substituting (36.9) into (36.8), we can cast the evolution of a wave function into a multiplicative form

$$\psi(q, t) = \int dq' K(q, q', t) \psi(q', 0),$$

with the kernel

$$K(q, q', t) = \sum_n \phi_n(q) e^{-iE_n t/\hbar} \phi_n^*(q') \quad (36.10)$$

called the *quantum evolution operator*, or the *propagator*. Applied twice, first for time t_1 and then for time t_2 , it propagates the initial wave function from q' to q'' , and then from q'' to q

$$K(q, q', t_1 + t_2) = \int dq'' K(q, q'', t_2) K(q'', q', t_1) \quad (36.11)$$

forward in time (hence the name ‘propagator’). In non-relativistic quantum mechanics, the range of q'' is infinite, so that the wave can propagate at any speed; in relativistic quantum mechanics, this is rectified by restricting the propagation to the forward light cone.

Because the propagator is a linear combination of the eigenfunctions of the Schrödinger equation, it too satisfies this equation

$$i\hbar \frac{\partial}{\partial t} K(q, q', t) = \hat{H}\left(q, \frac{i}{\hbar} \frac{\partial}{\partial q}\right) K(q, q', t), \quad (36.12)$$

and is thus a wave function defined for $t \geq 0$; from the completeness relation (36.6), we obtain the boundary condition at $t = 0$:

$$\lim_{t \rightarrow 0^+} K(q, q', t) = \delta(q - q'). \quad (36.13)$$

The propagator thus represents the time-evolution of a wave packet starting out as a configuration space delta-function localized at the point q' at initial time $t = 0$.

For time-independent Hamiltonians, the time dependence of the wave functions is known as soon as the eigenenergies E_n and eigenfunctions ϕ_n have been determined. With time dependence taken care of, it makes sense to focus on the *Green's function*, which is the Laplace transform of the propagator

$$G(q, q', E + i\epsilon) = \frac{1}{i\hbar} \int_0^\infty dt e^{\frac{i}{\hbar}Et - \frac{\epsilon}{\hbar}t} K(q, q', t) = \sum_n \frac{\phi_n(q)\phi_n^*(q')}{E - E_n + i\epsilon}. \quad (36.14)$$

Here, ϵ is a small positive number, ensuring the existence of the integral. The eigenenergies show up as poles in the Green's function with residues corresponding to the wave function amplitudes. If one is only interested in spectra, one may restrict oneself to the (formal) trace of the Green's function,

$$\text{Tr } G(q, q', E) = \int dq G(q, q, E) = \sum_n \frac{1}{E - E_n}, \quad (36.15)$$

where E is complex, with a positive imaginary part, and we have used the eigenfunction orthonormality (36.5). This trace is formal, because the sum in (36.15) is often divergent. We shall return to this point in sects. 39.1.1 and 39.1.2.

A useful characterization of the set of eigenvalues is given in terms of the *density of states*, with a delta function peak at each eigenenergy, figure 36.1 (a),

$$d(E) = \sum_n \delta(E - E_n). \quad (36.16)$$

Using the identity

$$\delta(E - E_n) = - \lim_{\epsilon \rightarrow +0} \frac{1}{\pi} \text{Im} \frac{1}{E - E_n + i\epsilon} \quad (36.17)$$

exercise 36.1

we can express the density of states in terms of the trace of the Green's function. That is,

$$d(E) = \sum_n \delta(E - E_n) = - \lim_{\epsilon \rightarrow 0} \frac{1}{\pi} \text{Im} \text{Tr } G(q, q', E + i\epsilon). \quad (36.18)$$

As we shall see (after “some” work), a semiclassical formula for the right-hand-side of this relation yields the quantum spectrum in terms of periodic orbits.

section 39.1.1

The density of states can be written as the derivative $d(E) = dN(E)/dE$ of the *spectral staircase* function

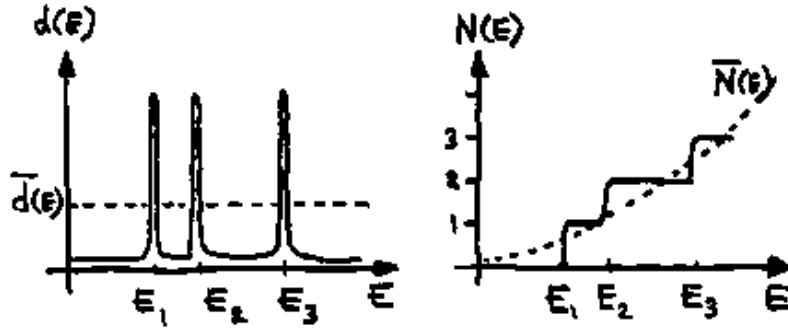
$$N(E) = \sum_n \Theta(E - E_n) \quad (36.19)$$

which counts the number of eigenenergies below E , figure 36.1 (b). Here Θ is the Heaviside function

$$\Theta(x) = 1 \quad \text{if } x > 0; \quad \Theta(x) = 0 \quad \text{if } x < 0. \quad (36.20)$$

The spectral staircase is a useful quantity in many contexts, both experimental and theoretical. This completes our lightning review of quantum mechanics.

Figure 36.1: Schematic picture of **a)** the density of states $d(E)$, and **b)** the spectral staircase function $N(E)$. The dashed lines denote the mean density of states $\bar{d}(E)$ and the average number of states $\bar{N}(E)$ discussed in more detail in sect. 39.1.1.



36.1 Exercises

36.1. **Dirac delta function, Lorentzian representation.** Derive the representation (36.17)

$$\delta(E - E_n) = - \lim_{\epsilon \rightarrow +0} \frac{1}{\pi} \text{Im} \frac{1}{E - E_n + i\epsilon} \quad (36.21)$$

of a delta function as imaginary part of $1/x$.

(Hint: read up on principal parts, positive and negative frequency part of the delta function, the Cauchy theorem in a good quantum mechanics textbook).

36.2. **Green's function.** Verify Green's function Laplace transform (36.14),

$$\begin{aligned} G(q, q', E + i\epsilon) &= \frac{1}{i\hbar} \int_0^\infty dt e^{\frac{i}{\hbar}Et - \frac{\epsilon}{\hbar}t} K(q, q', t) \\ &= \sum \frac{\phi_n(q)\phi_n^*(q')}{E - E_n + i\epsilon}, \end{aligned}$$

argue that positive ϵ is needed (hint: read a good quantum mechanics textbook).

36.3. **Scalar field propagator.** [M. Srednicki, Quantum Field Theory, Part I [arXiv:hep-th/0409035](https://arxiv.org/abs/hep-th/0409035), problem 8.2]

Starting with

$$\Delta(x - x') = \int \frac{d^4k}{(2\pi)^4} \frac{e^{ik(x-x')}}{k^2 + m^2 - i\epsilon}, \quad (36.22)$$

verify

$$\Delta(x - x') = i \int \tilde{d}k e^{ik \cdot (x-x') - i\omega|t-t'|} \quad (36.23)$$

$$\begin{aligned} &= i\theta(t - t') \int \tilde{d}k e^{ik(x-x')} \\ &\quad + i\theta(t' - t) \int \tilde{d}k e^{-ik(x-x')}. \end{aligned} \quad (36.24)$$

There should be an i in eq. (36.23).

36.4. **Quantum mechanical path integrals.**

Consider a particle moving in one dimension with separable Hamiltonian:

$$H = \frac{p^2}{2m} + V(q).$$

If the particle is at a position q at time $t = 0$, what is the probability amplitude that it will be at position q' at a later time $t = T$?

a) Show that the propagator from the initial spacetime point $(q, 0)$ to the final point (q', T) is

$$A = \langle q' | \psi(T) \rangle \equiv K(q', T; q, 0) = \langle q' | e^{-iHT} | q \rangle \quad (36.25)$$

(set $\hbar$ to 1 throughout).

b) Split the time evolution into two intervals $T = (T - t_1) + t_1$, show that the amplitude can be written as

$$A = \int dq_1 K(q', T; q_1, t_1) K(q_1, t_1; q, 0). \quad (36.26)$$

c) Split the time interval T into a large number N of short time intervals of duration $\epsilon = T/N$:

$$A = \langle q' | \left(e^{-iH\epsilon} \right)^N | q \rangle = \langle q' | \underbrace{e^{-iH\epsilon} e^{-iH\epsilon} \dots e^{-iH\epsilon}}_{N \text{ times}} | q \rangle,$$

and show that the amplitude can be written as the sum over the amplitudes of all possible N -segment paths,

$$A = \sum_{\text{paths}} A_{\text{path}},$$

where

$$\sum_{\text{paths}} = \int dq_1 \dots dq_{N-1},$$

$$A_{\text{path}} = K_{q_N, q_{N-1}} K_{q_{N-1}, q_{N-2}} \dots K_{q_2, q_1} K_{q_1, q_0}.$$

d) Show that the short-time propagator for one sub-interval is given by

$$K_{q_{j+1}, q_j} = \int \frac{dp_j}{2\pi} e^{ip_j(q_{j+1} - q_j)} e^{-i\epsilon H(p_j, \bar{q}_j)} (1 + o(\epsilon^2)), \quad (36.27)$$

where $\bar{q}_j = \frac{1}{2}(q_j + q_{j+1})$.

e) Combine N such factors, show that to the order of $o(\epsilon^2)$ the amplitude of a path is

$$A_{\text{path}} = \int \prod_{j=0}^{N-1} \frac{dp_j}{2\pi} \exp i\epsilon \sum_{j=0}^{N-1} (p_j \dot{q}_j - H(p_j, \bar{q}_j)). \quad (36.28)$$

f) Show that the propagator is given by

$$K = \int \prod_{j=1}^{N-1} dq_j \int \prod_{j=0}^{N-1} \frac{dp_j}{2\pi} \exp i\epsilon \sum_{j=0}^{N-1} (p_j \dot{q}_j - H(p_j, \bar{q}_j)), \quad (36.29)$$

with one momentum integral for each segment (N total), and one position integral for each intermediate position ($N - 1$ total). In the $N \rightarrow \infty$ limit, this is a *phase-space path integral* integral over all functions $p(t)$, $q(t)$:

$$K \equiv \int \mathcal{D}p(t) \mathcal{D}q(t) \exp i \int_0^T dt (p\dot{q} - H(p, q)), \quad (36.30)$$

with fixed ends $q(0) = q$, $q(T) = q'$.

g) Show that if the Hamiltonian is separable, $H = p^2/2m + V(q)$, one can carry out the momentum integrals in (36.29), and the propagator becomes

$$\begin{aligned} K &= \int \prod_{j=1}^{N-1} dq_j \exp -i\epsilon \sum_{j=0}^{N-1} V(\bar{q}_j) \prod_{j=0}^{N-1} \left(\sqrt{\frac{m}{2\pi i\epsilon}} \exp i\epsilon \frac{m\dot{q}_j^2}{2} \right) \\ &= \left(\frac{m}{2\pi i\epsilon} \right)^{N/2} \int \prod_{j=1}^{N-1} dq_j \exp i\epsilon \sum_{j=0}^{N-1} \left(\frac{m\dot{q}_j^2}{2} - V(\bar{q}_j) \right). \end{aligned} \quad (36.31)$$

The argument of the exponential is the action of a path passing through the points $q_0 = q, q_1, \dots, q_{N-1}, q_N = q'$. This is the *configuration space path integral*

$$K = \int \mathcal{D}q(t) e^{iS[q(t)]}. \quad (36.32)$$

- 36.5. **Yukawa equation.** For the $\mu^2 = -k^2 > 0$ (imaginary wavenumber k), the Helmholtz equation is known as the *screened Poisson equation* [1],

$$(-\nabla^2 + \mu^2) \phi(x) = 4\pi\rho(x), \quad x \in \mathbb{R}^d, \quad (36.33)$$

due to its applications to electric field screening in plasmas. In chemistry the equation governs steady-state diffusion in presence of a solute $\rho(x)$, or heat diffusion in presence of heat sources. In the relativistic quantum field theory, a free massive scalar boson in $D + 1$ dimensions is described by *Klein–Gordon* or *Yukawa equation*,

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2 + \frac{\mu^2 c^2}{\hbar^2} \right) \phi = 0. \quad (36.34)$$

Its steady, time-independent solutions reduce to the homogeneous screened Poisson equation, which we refer to as the D -dimensional Euclidean theory Klein–Gordon equation,

$$(-\nabla^2 + \mu^2) \phi(\mathbf{r}) = 0. \quad (36.35)$$

The solutions of the screened Poisson equation (36.33) are of the same form as for the Helmholtz equation, but with the oscillatory $\sin, \cos$, and $\exp(i \cdots)$ terms replaced by the hyperbolic $\sinh, \cosh$, and $\exp(- \cdots)$.

The Klein–Gordon Green’s function satisfies $g(x, x')$

$$(-\nabla^2 + \mu^2) g(x, x') = \delta(x - x'), \quad (36.36)$$

with the appropriate boundary conditions.

- (a) Argue on dimensional grounds, assuming isotropy and translational invariance, that the massless, $\mu = 0$ Euclidean Green’s function is proportional to

$$g(x, x') \propto \frac{1}{|x - x'|^{1-D/2}}, \quad (36.37)$$

The outgoing is $D = 3$ known as the *Yukawa potential*, the static, spherically symmetric solution

$$g(x, x') = -\frac{e^{-\mu|x-x'|}}{4\pi|x-x'|}. \quad (36.38)$$

to the Klein–Gordon equation. The Fourier transform relates the Yukawa potential to the massive scalar particle propagator, i.e., Green’s function of the static Klein–Gordon equation

$$V(\mathbf{r}) = \frac{-g^2}{(2\pi)^3} \int e^{i\mathbf{k}\cdot\mathbf{r}} \frac{4\pi}{k^2 + \mu^2} d^3k.$$

In $d = 2$ this integral can be explicitly evaluated as a Bessel function, Recheck the 2π factors

$$g(\mathbf{r}, 0) = \frac{1}{2\pi} \int_0^{+\infty} dk_r \frac{k_r J_0(k_r r)}{k_r^2 + \mu^2} = \frac{1}{2\pi} K_0(r\mu).$$

Chapter 37

WKB quantization

THE WAVE FUNCTION for a particle of energy E moving in a constant potential V is

$$\psi = Ae^{\frac{i}{\hbar}pq} \quad (37.1)$$

with a constant amplitude A , and constant wavelength $\lambda = 2\pi/k$, $k = p/\hbar$, and $p = \pm\sqrt{2m(E - V)}$ is the momentum. Here we generalize this solution to the case where the potential varies slowly over many wavelengths. This semiclassical (or WKB) approximate solution of the Schrödinger equation fails at deterministic turning points, configuration space points where the particle momentum vanishes. In such neighborhoods, where the semiclassical approximation fails, one needs to solve locally the exact quantum problem, in order to compute connection coefficients which patch up semiclassical segments into an approximate global wave function.

Two lessons follow. First, semiclassical methods can be very powerful - deterministic mechanics computations yield surprisingly accurate estimates of quantal spectra, without solving the Schrödinger equation. Second, semiclassical quantization does depend on a purely wave-mechanical phenomena, the coherent addition of phases accrued by all fixed energy phase space trajectories that connect pairs of coordinate points, and the topological phase loss at every turning point, a topological property of the deterministic flow that plays no role in classical mechanics.

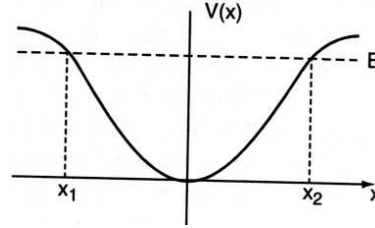
The semiclassical quantum field theory is sketched in appendix [A37](#).

37.1 WKB ansatz

If the kinetic term $T(p)$ can be separated as in [\(36.2\)](#), the time-independent Schrödinger equation takes form

$$-\frac{\hbar^2}{2m}\psi''(q) + V(q)\psi(q) = E\psi(q). \quad (37.2)$$

Figure 37.1: A 1-dimensional potential with two turning points at fixed energy E .



Consider a time-independent Schrödinger equation in 1 spatial dimension, with potential $V(q)$ growing sufficiently fast as $q \rightarrow \pm\infty$ so that the deterministic particle motion is confined for any E . Define the local momentum $p(q)$ and the local wavenumber $k(q)$ by

$$p(q) = \pm \sqrt{2m(E - V(q))}, \quad p(q) = \hbar k(q). \quad (37.3)$$

The variable wavenumber form of the Schrödinger equation

$$\psi'' + k^2(q)\psi = 0 \quad (37.4)$$

suggests that the wave function be written as $\psi = Ae^{\frac{i}{\hbar}S}$, A and S real functions of q . Substitution yields two equations, one for the real and other for the imaginary part:

$$(S')^2 = p^2 + \hbar^2 \frac{A''}{A} \quad (37.5)$$

$$S''A + 2S'A' = \frac{1}{A} \frac{d}{dq}(S'A^2) = 0. \quad (37.6)$$

The Wentzel-Kramers-Brillouin (*WKB*) or *semiclassical* approximation consists of dropping the $\hbar^2$ term in (37.5). Recalling that $p = \hbar k$, this amounts to assuming that $k^2 \gg \frac{A''}{A}$, which in turn implies that the phase of the wave function is changing much faster than its overall amplitude. So the WKB approximation can be interpreted either as a short wavelength/high frequency approximation to a wave-mechanical problem, or as the semiclassical, $\hbar \ll 1$ approximation to quantum mechanics.

Setting $\hbar = 0$ and integrating (37.5) we obtain the phase increment of a wave function initially at q , at energy E

$$S(q, q', E) = \int_{q'}^q dq'' p(q''). \quad (37.7)$$

This integral over a particle trajectory of constant energy, called the *action*, will play a key role in all that follows. The integration of (37.6) is even easier

$$A(q) = \frac{C}{|p(q)|^{\frac{1}{2}}}, \quad C = |p(q')|^{\frac{1}{2}} \psi(q'), \quad (37.8)$$

where the integration constant C is fixed by the value of the wave function at the initial point q' . The *WKB* (or *semiclassical*) *ansatz* wave function is given by

$$\psi_{sc}(q, q', E) = \frac{C}{|p(q)|^{\frac{1}{2}}} e^{\frac{i}{\hbar} S(q, q', E)}. \quad (37.9)$$

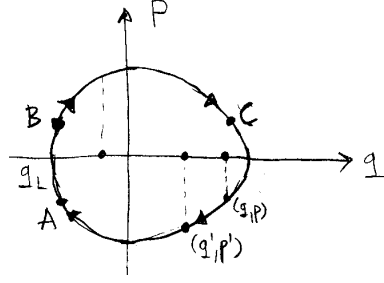


Figure 37.2: A 1-dof phase space trajectory of a particle moving in a bound potential.

In what follows we shall suppress dependence on the initial point and energy in such formulas, $(q, q', E) \rightarrow (q)$.

The WKB ansatz generalizes the free motion wave function (37.1), with the probability density $|A(q)|^2$ for finding a particle at q now inversely proportional to the velocity at that point, and the phase $\frac{1}{\hbar} q p$ replaced by $\frac{1}{\hbar} \int dq p(q)$, the integrated action along the trajectory. This is fine, except at any turning point q_0 , figure 37.1, where all energy is potential, and

$$p(q) \rightarrow 0 \quad \text{as} \quad q \rightarrow q_0, \quad (37.10)$$

so that the assumption that $k^2 \gg \frac{A''}{A}$ fails. What can one do in this case?

For the task at hand, a simple physical picture, due to Maslov, does the job. In the q coordinate, the turning points are defined by the zero kinetic energy condition (see figure 37.1), and the motion appears singular. This is not so in the full phase space: the trajectory in a smooth confining 1-dimensional potential is always a smooth loop (see figure 37.2), with the “special” role of the turning points q_L, q_R seen to be an artifact of a particular choice of the (q, p) coordinate frame. Maslov proceeds from the initial point (q', p') to a point (q_A, p_A) preceding the turning point in the $\psi(q)$ representation, then switch to the momentum representation

$$\tilde{\psi}(p) = \frac{1}{\sqrt{2\pi\hbar}} \int dq e^{-\frac{i}{\hbar} q p} \psi(q), \quad (37.11)$$

continue from (q_A, p_A) to (q_B, p_B) , switch back to the coordinate representation,

$$\psi(q) = \frac{1}{\sqrt{2\pi\hbar}} \int dp e^{\frac{i}{\hbar} q p} \tilde{\psi}(p), \quad (37.12)$$

and so on.

The only rub is that one usually cannot evaluate these transforms exactly. But, as the WKB wave function (37.9) is approximate anyway, it suffices to estimate these transforms to the leading order in $\hbar$ accuracy. This is accomplished by the method of stationary phase.

37.2 Method of stationary phase

All “semiclassical” approximations are based on saddle point evaluations of integrals of the type

$$I = \int dx A(x) e^{is\Phi(x)}, \quad x, \Phi(x) \in \mathbb{R}, \quad (37.13)$$

where s is a real parameter, and $\Phi(x)$ is a real-valued function. In our applications $s = 1/\hbar$ will always be assumed large.

For large s , the phase oscillates rapidly and “averages to zero” everywhere except at the *extremal points* $\Phi'(x_0) = 0$. The method of approximating an integral by its values at extremal points is called the *method of stationary phase*. Consider first the case of a 1-dimensional integral, and expand $\Phi(x_0 + \delta x)$ around x_0 to second order in δx ,

$$I = \int dx A(x) e^{is(\Phi(x_0) + \frac{1}{2}\Phi''(x_0)\delta x^2 + \dots)}. \quad (37.14)$$

Assume (for time being) that $\Phi''(x_0) \neq 0$, with either sign, $\text{sgn}[\Phi''] = \Phi''/|\Phi''| = \pm 1$. If in the neighborhood of x_0 the amplitude $A(x)$ varies slowly over many oscillations of the exponential function, we may retain the leading term in the Taylor expansion of the amplitude, and approximate the integral up to quadratic terms in the phase by

$$I \approx A(x_0) e^{is\Phi(x_0)} \int dx e^{\frac{1}{2}is\Phi''(x_0)(x-x_0)^2}. \quad (37.15)$$

The one integral that we know how to integrate is the Gaussian integral $\int dx e^{-\frac{x^2}{2b}} = \sqrt{2\pi b}$. For pure imaginary $b = ia$ one gets instead the *Fresnel integral formula*

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx e^{-\frac{x^2}{2ia}} = \sqrt{ia} = |a|^{1/2} e^{i\frac{\pi}{4} \frac{a}{|a|}} \quad (37.16)$$

exercise 37.2

we obtain

$$I \approx A(x_0) \left| \frac{2\pi}{s\Phi''(x_0)} \right|^{1/2} e^{is\Phi(x_0) \pm i\frac{\pi}{4}}, \quad (37.17)$$

where $\pm$ corresponds to the positive/negative sign of $s\Phi''(x_0)$.

37.3 WKB quantization

We can now evaluate the Fourier transforms (37.11), (37.12) to the same order in $\hbar$ as the WKB wave function using the stationary phase method,

$$\begin{aligned} \widetilde{\psi}_{sc}(p) &= \frac{C}{\sqrt{2\pi\hbar}} \int \frac{dq}{|p(q)|^{1/2}} e^{\frac{i}{\hbar}(S(q)-qp)} \\ &\approx \frac{C}{\sqrt{2\pi\hbar}} \frac{e^{\frac{i}{\hbar}(S(q^*)-q^*p)}}{|p(q^*)|^{1/2}} \int dq e^{\frac{i}{2\hbar}S''(q^*)(q-q^*)^2}, \end{aligned} \quad (37.18)$$

where q^* is given implicitly by the stationary phase condition

$$0 = S'(q^*) - p = p(q^*) - p$$

and the sign of $S''(q^*) = p'(q^*)$ determines the phase of the Fresnel integral (37.16)

$$\tilde{\psi}_{sc}(p) = \frac{C}{|p(q^*)p'(q^*)|^{\frac{1}{2}}} e^{\frac{i}{\hbar}[S(q^*) - q^*p] + \frac{i\pi}{4}\text{sgn}[S''(q^*)]}. \quad (37.19)$$

As we continue from (q_A, p_A) to (q_B, p_B) , nothing problematic occurs - $p(q^*)$ is finite, and so is the acceleration $p'(q^*)$. Otherwise, the trajectory would take infinitely long to get across. We recognize the exponent as the Legendre transform

$$\tilde{S}(p) = S(q(p)) - q(p)p$$

which can be used to express everything in terms of the p variable,

$$q^* = q(p), \quad \frac{d}{dq}q = 1 = \frac{dp}{dq} \frac{dq(p)}{dp} = q'(p)p'(q^*). \quad (37.20)$$

As the deterministic trajectory crosses q_L , the weight in (37.19),

$$\frac{d}{dq}p^2(q_L) = 2p(q_L)p'(q_L) = -2mV'(q_L), \quad (37.21)$$

is finite, and $S''(q^*) = p'(q^*) < 0$ for any point in the lower left quadrant, including (q_A, p_A) . Hence, the phase loss in (37.19) is $-\frac{\pi}{4}$. To go back from the p to the q representation, just turn figure 37.2 quarter-turn anticlockwise. Everything is the same if you replace $(q, p) \rightarrow (-p, q)$; so, without much ado we get the semiclassical wave function at the point (q_B, p_B) ,

$$\psi_{sc}(q) = \frac{e^{\frac{i}{\hbar}(\tilde{S}(p^*) + qp^*) - \frac{i\pi}{4}}}{|q^*(p^*)|^{\frac{1}{2}}} \tilde{\psi}_{sc}(p^*) = \frac{C}{|p(q)|^{\frac{1}{2}}} e^{\frac{i}{\hbar}S(q) - \frac{i\pi}{2}}. \quad (37.22)$$

The extra $|p'(q^*)|^{1/2}$ weight in (37.19) is cancelled by the $|q'(p^*)|^{1/2}$ term, by the Legendre relation (37.20).

The message is that going through a smooth potential turning point the WKB wave function phase slips by $-\frac{\pi}{2}$. This is equally true for the right and the left turning points, as can be seen by rotating figure 37.2 by 180° , and flipping coordinates $(q, p) \rightarrow (-q, -p)$. While a turning point is not an invariant concept (for a sufficiently short trajectory segment, it can be undone by a 45° turn), for a complete period $(q, p) = (q', p')$ the total phase slip is always $-2 \cdot \pi/2$, as a loop always has $m = 2$ turning points.

The *WKB quantization condition* follows by demanding that the wave function computed after a complete period be single-valued. With the normalization (37.8), we obtain

$$\psi(q') = \psi(q) = \left| \frac{p(q')}{p(q)} \right|^{\frac{1}{2}} e^{i(\frac{1}{\hbar} \oint p(q) dq - \pi)} \psi(q').$$

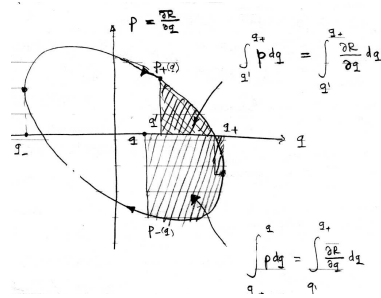


Figure 37.3: $S_p(E)$, the action of a periodic orbit p at energy E , equals the area in the phase space traced out by the 1-dof trajectory.

The prefactor is 1 by the periodic orbit condition $q = q'$, so the phase must be a multiple of 2π ,

$$\frac{1}{\hbar} \oint p(q) dq = 2\pi \left(n + \frac{m}{4} \right), \quad (37.23)$$

where m is the number of turning points along the trajectory - for this 1-dof problem, $m = 2$.

The action integral in (37.23) is the area (see figure 37.3) enclosed by the classical phase space loop of figure 37.2, and the quantization condition says that eigen-energies correspond to loops whose action is an integer multiple of the unit quantum of action, Planck's constant $\hbar$. The extra topological phase, which, although it had been discovered many times in centuries past, had to wait for its quantum chaotic (re)birth until the 1970's. Despite its derivation in a noninvariant coordinate frame, the final result involves only canonically invariant deterministic quantities, the periodic orbit action S , and the topological index m .

37.3.1 Harmonic oscillator quantization

Let us check the WKB quantization for one case (the only case?) whose quantum mechanics we fully understand: the harmonic oscillator

$$E = \frac{1}{2m} (p^2 + (m\omega q)^2).$$

The loop in figure 37.2 is now a circle in the $(m\omega q, p)$ plane, the action is its area $S = 2\pi E/\omega$, and the spectrum in the WKB approximation

$$E_n = \hbar\omega(n + 1/2) \quad (37.24)$$

turns out to be the *exact* harmonic oscillator spectrum. The stationary phase condition (37.18) keeps $V(q)$ accurate to order q^2 , which in this case is the whole answer (but we were simply lucky, really). For many 1-dof problems the WKB spectrum turns out to be very accurate all the way down to the ground state. Surprisingly accurate, if one interprets dropping the $\hbar^2$ term in (37.5) as a short wavelength approximation.

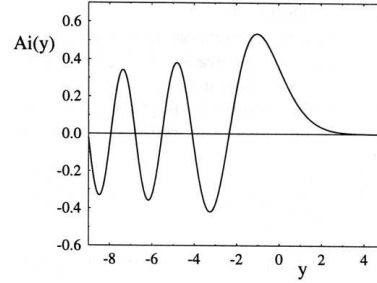


Figure 37.4: Airy function $Ai(q)$.

37.4 Beyond the quadratic saddle point

We showed, with a bit of Fresnel/Maslov voodoo, that in a smoothly varying potential the phase of the WKB wave function slips by a $\pi/2$ for each turning point. This $\pi/2$ came from a $\sqrt{i}$ in the Fresnel integral (37.16), one such factor for every time we switched representation from the configuration space to the momentum space, or back. Good, but what does this mean?

The stationary phase approximation (37.14) fails whenever $\Phi''(x) = 0$, or, in our the WKB ansatz (37.18), whenever the momentum $p'(q) = S''(q)$ vanishes. In that case we have to go beyond the quadratic approximation (37.15) to the first nonvanishing term in the Taylor expansion of the exponent. If $\Phi'''(x_0) \neq 0$, then

$$I \approx A(x_0) e^{is\Phi(x_0)} \int_{-\infty}^{\infty} dx e^{is\Phi'''(x_0) \frac{(x-x_0)^3}{6}}. \quad (37.25)$$

Airy functions can be represented by integrals of the form

$$Ai(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dy e^{i(xy - \frac{y^3}{3})}. \quad (37.26)$$

With a bit of Fresnel/Maslov voodoo we have shown that at each turning point a WKB wave function loses a bit of phase. Derivations of the WKB quantization condition given in standard quantum mechanics textbooks rely on expanding the potential close to the turning point

$$V(q) = V(q_0) + (q - q_0)V'(q_0) + \cdots,$$

solving the Airy equation (with $V'(q_0) \rightarrow z$ after appropriate rescalings),

$$\psi'' = z\psi, \quad (37.27)$$

and matching the oscillatory and the exponentially decaying “forbidden” region wave function pieces by means of the *WKB connection formulas*. That requires staring at Airy functions (see (37.4)) and learning about their asymptotics - a challenge that we will have to eventually overcome, in order to incorporate diffraction phenomena into semiclassical quantization.

The physical origin of the topological phase is illustrated by the shape of the Airy function, figure 37.4. For a potential with a finite slope $V'(q)$ the wave function penetrates into the forbidden region, and accommodates a bit more of a stationary wavelength than what one would expect from the deterministic trajectory alone. For infinite walls (i.e., billiards) a different argument applies: the wave function must vanish at the wall, and the phase slip due to a specular reflection is $-\pi$, rather than $-\pi/2$.

Résumé

The WKB ansatz wave function for 1-degree of freedom problems fails at the turning points of the deterministic trajectory. While in the q -representation the WKB ansatz at a turning point is singular, along the p direction the deterministic trajectory in the same neighborhood is smooth, as for any smooth bound potential the classical motion is topologically a circle around the origin in the (q, p) space. The simplest way to deal with such singularities is as follows; follow the deterministic trajectory in q -space until the WKB approximation fails close to the turning point; then insert $\int dp|p\rangle\langle p|$ and follow the deterministic trajectory in the p -space until you encounter the next p -space turning point; go back to the q -space representation, and so on. Each matching involves a Fresnel integral, yielding an extra $e^{-i\pi/4}$ phase shift, for a total of $e^{-i\pi}$ phase shift for a full period of a semi-classical particle moving in a soft potential. The condition that the wave-function be single-valued then leads to the 1-dimensional WKB quantization, and its lucky cousin, the Bohr-Sommerfeld quantization.

Alternatively, one can linearize the potential around the turning point a , $V(q) = V(a) + (q - a)V'(a) + \dots$, and solve the quantum mechanical constant linear potential $V(q) = qF$ problem exactly, in terms of an Airy function. An approximate wave function is then patched together from an Airy function at each turning point, and the WKB ansatz wave-function segments in-between via the WKB connection formulas. The single-valuedness condition again yields the 1-dimensional WKB quantization. This a bit more work than tracking the deterministic trajectory in the full phase space, but it gives us a better feeling for shapes of quantum eigenfunctions, and exemplifies the general strategy for dealing with other singularities, such as wedges, bifurcation points, creeping and tunneling: patch together the WKB segments by means of exact QM solutions to local approximations to singular points.

Commentary

Remark 37.1. Airy function. The stationary phase approximation is all that is needed for the semiclassical approximation, with the proviso that D in (38.36) has no zero eigenvalues. The zero eigenvalue case would require going beyond the Gaussian saddle-point approximation, which typically leads to approximations of the integrals in terms of Airy

functions [1]. S. Strogatz explains that clearly in  Lect. 20 *WKB for eigenvalue problems*, the *Asymptotics and perturbation methods* course (2021).

exercise 33.5

Remark 37.2. Bohr-Sommerfeld quantization. Bohr-Sommerfeld quantization condition [3] was the key result of the old quantum theory, in which the electron trajectories were purely deterministically. They were lucky - the symmetries of the Kepler problem work out in such a way that the total topological index $m = 4$ amount effectively to numbering the energy levels starting with $n = 1$. They were unlucky - because the hydrogen $m = 4$ masked the topological index, they could never get the helium spectrum right - the semiclassical calculation had to wait for until 1980, when Leopold and Percival [2] added the topological indices.

References

- [1] N. Bleistein and R. A. Handelsman, *Asymptotic Expansions of Integrals* (Dover, New York, 1986).
- [2] J. G. Leopold and I. Percival, “The semiclassical two-electron atom and the old quantum theory”, *J. Phys. B* **13**, 1037 (1980).
- [3] A. Sommerfeld, *Atombau und Spektrallinien* (Deutsch, Thun, Frankfurt/M, 1919).

Exercises

37.1. **WKB ansatz.**  Try to show that no other ansatz other than (38.1) gives a meaningful definition of the momentum in the $\hbar \rightarrow 0$ limit.

37.2. **Fresnel integral.** Derive the Fresnel integral

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx e^{-\frac{x^2}{2ia}} = \sqrt{ia} = |a|^{1/2} e^{i\frac{\pi}{4} \frac{a}{|a|}}.$$

37.3. **Sterling formula for $n!$.** Compute an approximate value of $n!$ for large n using the stationary phase approximation. Hint: $n! = \int_0^{\infty} dt t^n e^{-t}$.

Chapter 38

Semiclassical evolution

William Rowan Hamilton was born in 1805. At three he could read English; by four he began to read Latin, Greek and Hebrew, by ten he read Sanskrit, Persian, Arabic, Chaldee, Syrian and sundry Indian dialects. At age seventeen he began to think about optics, and worked out his great principle of “Characteristic Function.”

— Turnbull, *Lives of Mathematicians*

(G. Vattay, G. Tanner and P. Cvitanović)

SEMICLASSICAL APPROXIMATIONS to quantum mechanics are valid in the regime where the de Broglie wavelength $\lambda \sim \hbar/p$ of a particle with momentum p is much shorter than the length scales across which the potential of the system changes significantly. In the short wavelength approximation the particle is a point-like object bouncing off potential walls, the same way it does in the deterministic mechanics. The novelty of quantum mechanics is the interference of the point-like particle with other versions of itself traveling along different deterministic trajectories, a feat impossible in deterministic mechanics. The short wavelength – or semiclassical – formalism is developed by formally taking the limit $\hbar \rightarrow 0$ in quantum mechanics in such a way that quantum quantities go to their deterministic counterparts.

remark 38.1

38.1 Hamilton-Jacobi theory

We saw in chapter 37 that for a 1-degree of freedom particle moving in a slowly varying potential, it makes sense to generalize the free particle wave function (37.1) to a wave function

$$\psi(q, t) = A(q, t)e^{iR(q, t)/\hbar}, \quad (38.1)$$

with slowly varying (real) amplitude $A(q, t)$ and rapidly varying (real) phase $R(q, t)$. Its phase and magnitude. The time evolution of the phase and the magnitude of ψ

exercise 37.1

follows from the Schrödinger equation (36.1)

$$\left(i\hbar \frac{\partial}{\partial t} + \frac{\hbar^2}{2m} \frac{\partial^2}{\partial q^2} - V(q) \right) \psi(q, t) = 0. \quad (38.2)$$

Assume $A \neq 0$, and separate out the real and the imaginary parts. We get two equations: The real part governs the time evolution of the phase

$$\frac{\partial R}{\partial t} + \frac{1}{2m} \left(\frac{\partial R}{\partial q} \right)^2 + V(q) - \frac{\hbar^2}{2m} \frac{1}{A} \frac{\partial^2}{\partial q^2} A = 0, \quad (38.3)$$

and the imaginary part the time evolution of the amplitude

$$\frac{\partial A}{\partial t} + \frac{1}{m} \sum_{i=1}^D \frac{\partial A}{\partial q_i} \frac{\partial R}{\partial q_i} + \frac{1}{2m} A \frac{\partial^2 R}{\partial q^2} = 0. \quad (38.4)$$

exercise 38.6

exercise 38.7

exercise 38.8

In this way a linear PDE for a complex wave function is converted into a set of coupled non-linear PDE's for real-valued functions R and A . The coupling term in (38.3) is, however, of order $\hbar^2$ and thus small in the semiclassical limit $\hbar \rightarrow 0$.

Now we generalize the *Wentzel-Kramers-Brillouin* (WKB) *ansatz* for 1-degree of freedom dynamics to the *Van Vleck ansatz* in arbitrary dimension: we assume the magnitude $A(q, t)$ varies slowly compared to the phase $R(q, t)/\hbar$, so we drop the $\hbar$ -dependent term. In this approximation the phase $R(q, t)$ and the corresponding “momentum field” $\frac{\partial R}{\partial q}(q, t)$ can be determined from the amplitude independent equation

$$\frac{\partial R}{\partial t} + H\left(q, \frac{\partial R}{\partial q}\right) = 0. \quad (38.5)$$

In classical mechanics this equation is known as the *Hamilton-Jacobi equation*. We will refer to this step (as well as all leading order in $\hbar$ approximations to follow) as the *semiclassical approximation* to wave mechanics, and from now on work only within this approximation.

38.1.1 Hamilton's equations

We now solve the nonlinear partial differential equation (38.5) in a way the 17 year old Hamilton might have solved it. The main step is the step leading from the nonlinear PDE (38.9) to Hamilton's ODEs (38.10). If you already understand the Hamilton-Jacobi theory, you can safely skip this section.

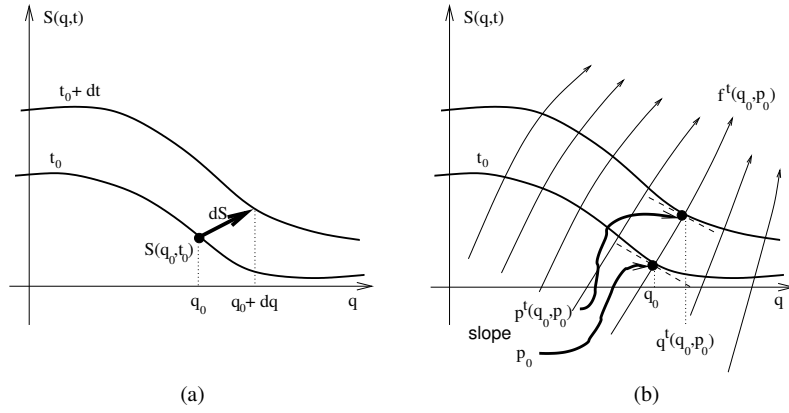


fast track:

sect. 38.1.3, p. 740

The wave equation (36.1) describes how the wave function ψ evolves with time, and if you think of ψ as an (infinite dimensional) vector, position q plays a

Figure 38.1: (a) A phase $R(q, t)$ plotted as a function of the position q for two infinitesimally close times. (b) The phase $R(q, t)$ transported by a swarm of “particles”. The Hamilton’s equations (38.10) construct $R(q, t)$ by transporting $q_0 \rightarrow q(t)$ and the slope of $R(q_0, t_0)$, that is $p_0 \rightarrow p(t)$.



role of an index. In one spatial dimension the phase R plotted as a function of the position q for two different times looks something like figure 38.1 (a): The phase $R(q, t_0)$ deforms smoothly with time into the phase $R(q, t)$ at time t . Hamilton’s idea was to let a swarm of particles transport R and its slope $\partial R/\partial q$ at q at initial time $t = t_0$ to a corresponding $R(q, t)$ and its slope at time t , figure 38.1 (b). For notational convenience, define

$$p_i = p_i(q, t) := \frac{\partial R}{\partial q_i}, \quad i = 1, 2, \dots, D. \quad (38.6)$$

We saw earlier that (38.3) reduces in the semiclassical approximation to the Hamilton-Jacobi equation (38.5). To make life simple, we shall assume throughout this chapter that the Hamilton’s function $H(q, p)$ does not depend explicitly on time t , i.e., the energy is conserved.

To start with, we also assume that the function $R(q, t)$ is smooth and well defined for every q at the initial time t . This is true for sufficiently short times; as we will see later, R develops folds and becomes multi-valued as t progresses. Consider now the variation of the function $R(q, t)$ with respect to independent infinitesimal variations of the time and space coordinates dt and dq , figure 38.1 (a)

$$dR = \frac{\partial R}{\partial t} dt + \frac{\partial R}{\partial q} dq. \quad (38.7)$$

Dividing through by dt and substituting (38.5) we obtain the total derivative of $R(q, t)$ with respect to time *along the as yet arbitrary direction* $\dot{q}$, that is,

$$\frac{dR}{dt}(q, \dot{q}, t) = -H(q, p) + \dot{q} \cdot p. \quad (38.8)$$

Note that the “momentum” $p = \partial R/\partial q$ is a well defined function of q and t . In order to integrate $R(q, t)$ with the help of (38.8) we also need to know how $p = \partial R/\partial q$ changes along $\dot{q}$. Varying p with respect to independent infinitesimal variations dt and dq and substituting the Hamilton-Jacobi equation (38.5) yields

$$d \frac{\partial R}{\partial q} = \frac{\partial^2 R}{\partial q \partial t} dt + \frac{\partial^2 R}{\partial q^2} dq = - \left(\frac{\partial H}{\partial q} + \frac{\partial H}{\partial p} \frac{\partial p}{\partial q} \right) dt + \frac{\partial p}{\partial q} dq.$$

Note that $H(q, p)$ depends on q also through $p(q, t) = \partial R / \partial q$, hence the $\frac{\partial H}{\partial p}$ term in the above equation. Dividing again through by dt we get the time derivative of $\partial R / \partial q$, that is,

$$\dot{p}(q, \dot{q}, t) + \frac{\partial H}{\partial q} = \left(\dot{q} - \frac{\partial H}{\partial p} \right) \frac{\partial p}{\partial q}. \quad (38.9)$$

Time variation of p depends not only on the yet unknown $\dot{q}$, but also on the second derivatives of R with respect to q with yet unknown time dependence. However, if we *choose* $\dot{q}$ (which was arbitrary, so far) such that the right hand side of the above equation vanishes, we can calculate the function $R(q, t)$ along a specific trajectory $(q(t), p(t))$ given by integrating the ordinary differential equations

$$\dot{q} = \frac{\partial H(q, p)}{\partial p}, \quad \dot{p} = -\frac{\partial H(q, p)}{\partial q} \quad (38.10)$$

with initial conditions

$$q(t_0) = q', \quad p(t_0) = p' = \frac{\partial R}{\partial q}(q', t_0). \quad (38.11)$$

We recognize (38.10) as Hamilton's equations of motion of classical mechanics. The miracle happens in the step leading from (38.5) to (38.9) – if you missed it, you have missed the point. Hamilton derived his equations contemplating optics - it took him three more years to realize that all of Newtonian dynamics can be profitably recast in this form.

section 8.1

$\dot{q}$ is no longer an independent function, and the phase $R(q, t)$ can now be computed by integrating equation (38.8) along the trajectory $(q(t), p(t))$

$$\begin{aligned} R(q, t) &= R(q', t_0) + R(q, t; q', t_0) \\ R(q, t; q', t_0) &= \int_{t_0}^t d\tau [\dot{q}(\tau) \cdot p(\tau) - H(q(\tau), p(\tau))] , \end{aligned} \quad (38.12)$$

with the initial conditions (38.11). In this way the Hamilton-Jacobi *partial* differential equation (38.3) is solved by integrating a set of *ordinary* differential equations, Hamilton's equations. In order to determine $R(q, t)$ for arbitrary q and t we have to find a q' such that the trajectory starting in $(q', p' = \partial_q R(q', t_0))$ reaches q in time t and then compute R along this trajectory, see figure 38.1 (b). The integrand of (38.12) is known as the *Lagrangian*,

$$L(q, \dot{q}, t) = \dot{q} \cdot p - H(q, p, t). \quad (38.13)$$

A variational principle lurks here, but we shall not make much fuss about it as yet.

Throughout this chapter we assume that the energy is conserved, and that the only time dependence of $H(q, p)$ is through $(q(\tau), p(\tau))$, so the value of $R(q, t; q', t_0)$ does not depend on t_0 , but only on the elapsed time $t - t_0$. To simplify notation we will set $t_0 = 0$ and write

$$R(q, q', t) = R(q, t; q', 0).$$

The initial momentum of the particle must coincide with the initial momentum of the trajectory connecting q' and q :

$$p' = \frac{\partial}{\partial q'} R(q', 0) = -\frac{\partial}{\partial q'} R(q, q', t). \quad (38.14)$$

The function $R(q, q', t)$ is known as *Hamilton's principal function*.

exercise 38.5
exercise 38.9

To summarize: Hamilton's achievement was to trade in the Hamilton-Jacobi *partial* differential equation (38.5) describing the evolution of a wave front for a finite number of *ordinary* differential equations of motion, with the initial phase $R(q, 0)$ incremented by the integral (38.12) evaluated along the phase space trajectory $(q(\tau), p(\tau))$.

38.1.2 Action

Before proceeding, we note in passing a few facts about Hamiltonian dynamics that will be needed for the construction of semiclassical Green's functions. If the energy is conserved, the $\int H(q, p) d\tau$ integral in (38.12) is simply Et . The first term, or the *action*

$$S(q, q', E) = \int_0^t d\tau \dot{q}(\tau) \cdot p(\tau) = \int_{q'}^q dq \cdot p \quad (38.15)$$

is integrated along a trajectory from q' to q with a fixed energy E . By (38.12) the action is a Legendre transform of Hamilton's principal function

$$S(q, q', E) = R(q, q', t) + Et. \quad (38.16)$$

The time of flight t along the trajectory connecting $q' \rightarrow q$ with fixed energy E is given by

$$\frac{\partial}{\partial E} S(q, q', E) = t. \quad (38.17)$$

The way to think about the formula (38.16) for action is that the time of flight is a function of the energy, $t = t(q, q', E)$. The left hand side is explicitly a function of E ; the right hand side is an implicit function of E through energy dependence of the flight time t .

Going in the opposite direction, the energy of a trajectory $E = E(q, q', t)$ connecting $q' \rightarrow q$ with a given time of flight t is given by the derivative of Hamilton's principal function

$$\frac{\partial}{\partial t} R(q, q', t) = -E, \quad (38.18)$$

and the second variations of R and S are related in the standard way of Legendre transforms:

$$\frac{\partial^2}{\partial t^2} R(q, q', t) \frac{\partial^2}{\partial E^2} S(q, q', E) = -1. \quad (38.19)$$

A geometric visualization of what the phase evolution looks like is very helpful in understanding the origin of topological indices to be introduced in what follows. Given an initial phase $R(q, t_0)$, the gradient $\partial_q R$ defines a D -dimensional *Lagrangian manifold* ($q, p = \partial_q R(q)$) in the full $2D$ dimensional phase space (q, p) . The defining property of this manifold is that any contractible loop γ in it has zero action,

section 38.1.4

$$0 = \oint_{\gamma} dq \cdot p,$$

a fact that follows from the definition of p as a gradient, and the Stokes theorem. Hamilton's equations of motion preserve this property and map a Lagrangian manifold into a Lagrangian manifold at a later time. t

Returning back to the main line of our argument: so far we have determined the wave function phase $R(q, t)$. Next we show that the velocity field given by the Hamilton's equations together with the continuity equation determines the amplitude of the wave function.

38.1.3 Density evolution

To obtain the full solution of the Schrödinger equation (36.1), we also have to integrate (38.4).

$$\rho(q, t) := A^2 = \psi^* \psi$$

plays the role of a density. To the leading order in $\hbar$, the gradient of R may be interpreted as the semiclassical momentum density

$$\psi(q, t)^* (-i\hbar \frac{\partial}{\partial q}) \psi(q, t) = -i\hbar A \frac{\partial A}{\partial q} + \rho \frac{\partial R}{\partial q}.$$

Evaluated along the trajectory $(q(t), p(t))$, the amplitude equation (38.4) is equivalent to the continuity equation (19.33) after multiplying (38.4) by $2A$, that is

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial q_i} (\rho v_i) = 0. \quad (38.20)$$

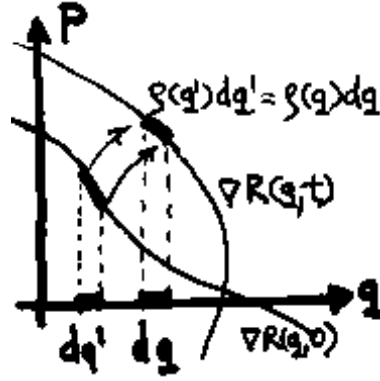
Here, $v_i = \dot{q}_i = p_i/m$ denotes a velocity field, which is in turn determined by the gradient of $R(q, t)$, or the *Lagrangian manifold* $(q(t), p(t) = \partial_q R(q, t))$,

$$v = \frac{1}{m} \frac{\partial}{\partial q} R(q, t).$$

As we already know how to solve the Hamilton-Jacobi equation (38.5), we can also solve for the density evolution as follows:

The density $\rho(q)$ can be visualized as the density of a configuration space flow $q(t)$ of a swarm of hypothetical particles; the trajectories $q(t)$ are solutions of Hamilton's equations with initial conditions given by $(q(0) = q', p(0) = p' = \partial_q R(q', 0))$.

Figure 38.2: Density evolution of an initial surface $(q', p' = \partial_{q'} R(q', 0))$ into $(q(t), p(t))$ surface time t later, sketched in 1 dimension. While the number of trajectories and the phase space Liouville volume are conserved, the density of trajectories projected on the q coordinate varies; trajectories which started in dq' at time zero end up in the interval dq .



If we take a small configuration space volume $d^D q$ around some point q at time t , then the number of particles in it is $\rho(q, t)d^D q$. They started initially in a small volume $d^D q'$ around the point q' of the configuration space. For the moment, we assume that there is only one solution, the case of several paths will be considered below. The number of particles at time t in the volume is the same as the number of particles in the initial volume at $t = 0$,

$$\rho(q(t), t)d^D q = \rho(q', 0)d^D q',$$

see figure 38.2. The ratio of the initial and the final volumes can be expressed as

$$\rho(q(t), t) = \left| \det \frac{\partial q'}{\partial q} \right| \rho(q', 0). \quad (38.21)$$

As we know how to compute trajectories $(q(t), p(t))$, we know how to compute this Jacobian and, by (38.21), the density $\rho(q(t), t)$ at time t .

section 19.2

38.1.4 Semiclassical wave function

Now we have all ingredients to write down the semiclassical wave function at time t . Consider first the case when our initial wave function can be written in terms of single-valued functions $A(q', 0)$ and $R(q', 0)$. For sufficiently short times, $R(q, t)$ will remain a single-valued function of q , and every $d^D q$ configuration space volume element keeps its orientation. The evolved wave function is in the semiclassical approximation then given by

$$\begin{aligned} \psi_{sc}(q, t) &= A(q, t)e^{iR(q, t)/\hbar} = \sqrt{\det \frac{\partial q'}{\partial q}} A(q', 0)e^{i(R(q', 0) + R(q, q', t))/\hbar} \\ &= \sqrt{\det \frac{\partial q'}{\partial q}} e^{iR(q, q', t)/\hbar} \psi(q', 0). \end{aligned}$$

As the time progresses the Lagrangian manifold $\partial_q R(q, t)$ can develop folds, so for longer times the value of the phase $R(q, t)$ is not necessarily unique; in general more than one trajectory will connect points q and q' with different phases $R(q, q', t)$ accumulated along these paths, see figure 38.3.

We thus expect in general a collection of different trajectories from q' to q which we will index by j , with different phase increments $R_j(q, q', t)$. The hypothetical particles of the density flow at a given configuration space point can move with different momenta $p = \partial_q R_j(q, t)$. This is not an ambiguity, since in the full (q, p) phase space each particle follows its own trajectory with a unique momentum.

Whenever the Lagrangian manifold develops a fold, the density of the phase space trajectories in the fold projected on the configuration coordinates diverges. As illustrated in figure 38.3, when the Lagrangian manifold develops a fold at $q = q_1$; the volume element dq_1 in the neighborhood of the folding point is proportional to $\sqrt{dq'}$ instead of dq' . The Jacobian $\partial q' / \partial q$ diverges like $1 / \sqrt{q_1 - q(t)}$ when computed along the trajectory going through the folding point at q_1 . After the folding the orientation of the interval dq' has changed when being mapped into dq_2 ; in addition the function R , as well as its derivative which defines the Lagrangian manifold, becomes multi-valued. Distinct trajectories starting from different initial points q' can now reach the same final point q_2 . (That is, the point q' may have more than one pre-image.) The projection of a simple fold, or of an envelope of a family of phase space trajectories, is called a *caustic*; this expression comes from the Greek word for “capable of burning,” evoking the luminous patterns that one observes swirling across the bottom of a swimming pool.

The folding also changes the orientation of the pieces of the Lagrangian manifold $(q, \partial_q R(q, t))$ with respect to the initial manifold, so the eigenvalues of the Jacobian determinant change sign at each fold crossing. We can keep track of the signs by writing the Jacobian determinant as

$$\det \frac{\partial q'}{\partial q} \Big|_j = e^{-i\pi m_j(q, q', t)} \left| \det \frac{\partial q'}{\partial q} \right|_j,$$

where $m_j(q, q', t)$ counts the number of sign changes of the Jacobian determinant on the way from q' to q along the trajectory indexed with j , see figure 38.3. We shall refer to the integer $m_j(q, q', t)$ as the *topological* of the trajectory. So in general the semiclassical approximation to the wave function is thus a sum over possible trajectories that start at any initial q' and end in q in time t

$$\psi_{sc}(q, t) = \int dq' \sum_j \left| \det \frac{\partial q'}{\partial q} \right|_j^{1/2} e^{iR_j(q, q', t)/\hbar - i\pi m_j(q, q', t)/2} \psi(q'_j, 0), \quad (38.22)$$

each contribution weighted by corresponding density, phase increment and the topological index.

That the correct topological index is obtained by simply counting the number of eigenvalue sign changes and taking the square root is not obvious - the careful argument requires that quantum wave functions evaluated across the folds remain single valued.

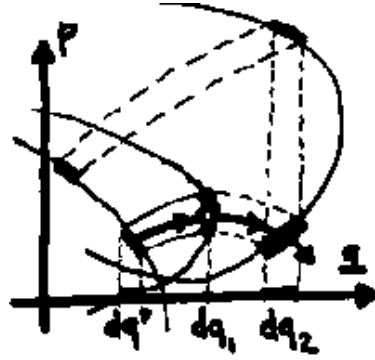


Figure 38.3: Folding of the Lagrangian surface $(q, \partial_q R(q, t))$.

38.2 Semiclassical propagator

We saw in chapter 36 that the evolution of an initial wave function $\psi(q, 0)$ is completely determined by the propagator (36.10). As $K(q, q', t)$ itself satisfies the Schrödinger equation (36.12), we can treat it as a wave function parameterized by the configuration point q' . In order to obtain a semiclassical approximation to the propagator we follow now the ideas developed in the last section. There is, however, one small complication: the initial condition (36.13) demands that the propagator at $t = 0$ is a δ -function at $q = q'$, that is, the amplitude is infinite at $q = q'$ and the phase is not well defined. Our hypothetical cloud of particles is thus initially localized at $q = q'$ with *any* initial velocity. This is in contrast to the situation in the previous section where we assumed that the particles at a given point q have well defined velocity (or a discrete set of velocities) given by $\dot{q} = \partial_p H(q, p)$. We will now derive a semiclassical expression for $K(q, q', t)$ by considering the propagator for short times first, and extrapolating from there to arbitrary times t .

38.2.1 Short time propagator

For infinitesimally short times δt away from the singular point $t = 0$ we assume that it is again possible to write the propagator in terms of a well defined phase and amplitude, that is

$$K(q, q', \delta t) = A(q, q', \delta t) e^{\frac{i}{\hbar} R(q, q', \delta t)}.$$

As all particles start at $q = q'$, $R(q, q', \delta t)$ will be of the form (38.12), that is

$$R(q, q', \delta t) = p \dot{q} \delta t - H(q, p) \delta t, \quad (38.23)$$

with $\dot{q} \approx (q - q')/\delta t$. For Hamiltonians of the form (36.2) we have $\dot{q} = p/m$, which leads to

$$R(q, q', \delta t) = \frac{m(q - q')^2}{2 \delta t} - V(q) \delta t.$$

Here V can be evaluated any place along the trajectory from q to q' , for example at the midway point $V((q + q')/2)$. Inserting this into our ansatz for the propagator we obtain

$$K_{sc}(q, q', \delta t) \approx A(q, q', \delta t) e^{\frac{i}{\hbar} \left(\frac{m}{2 \delta t} (q - q')^2 - V(q) \delta t \right)}. \quad (38.24)$$

For infinitesimal times we can neglect the term $V(q)\delta t$, so $K_{sc}(q, q', \delta t)$ is a d -dimensional Gaussian with width $\sigma^2 = i\hbar\delta t/m$. This Gaussian is a finite width approximation to the Dirac delta function

$$\delta(z) = \lim_{\sigma \rightarrow 0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2\sigma^2} \quad (38.25)$$

if $A = (m/2\pi i\hbar\delta t)^{D/2}$, with $A(q, q', \delta t)$ fixed by the Dirac delta function normalization condition. The correctly normalized propagator for infinitesimal times δt is therefore

exercise 38.1

$$K_{sc}(q, q', \delta t) \approx \left(\frac{m}{2\pi i\hbar\delta t} \right)^{D/2} e^{i\left(\frac{m(q-q')^2}{2\delta t} - V(q)\delta t \right)}. \quad (38.26)$$

The short time dynamics of the Lagrangian manifold $(q, \partial_q R)$ which corresponds to the quantum propagator can now be deduced from (38.23); one obtains

$$\frac{\partial R}{\partial q} = p \approx \frac{m}{\delta t}(q - q').$$

Note, that for $\delta t \rightarrow 0$, this plane is given by the condition $q = q'$, that is, particles start on a plane parallel to the momentum axis. As we have already noted, all particles start at $q = q'$ but with different velocities for $t = 0$. The initial surface $(q', p' = \partial_q R(q', 0))$ is mapped into the surface $(q(t), p(t))$ some time t later. The slope of the Lagrangian plane for a short finite time is given as

$$\frac{\partial p_i}{\partial q_j} = -\frac{\partial^2 R}{\partial q_j \partial q'_i} = -\frac{\partial p'_i}{\partial q_j} = \frac{m}{\delta t} \delta_{ij}.$$

The prefactor $(m/\delta t)^{D/2}$ in (38.26) can therefore be interpreted as the determinant of the Jacobian of the transformation from final position coordinates q to initial momentum coordinates p' , that is

$$K_{sc}(q, q', \delta t) = \frac{1}{(2\pi i\hbar)^{D/2}} \left(\det \frac{\partial p'_i}{\partial q_j} \right)^{1/2} e^{iR(q, q', \delta t)/\hbar}, \quad (38.27)$$

where

$$\left. \frac{\partial p'_i}{\partial q_j} \right|_{t, q'} = \frac{\partial^2 R(q, q', \delta t)}{\partial q_j \partial q'_i} \quad (38.28)$$

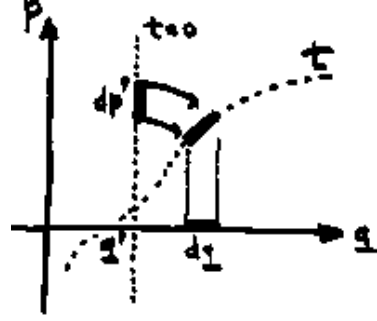
The subscript $\cdots|_{t, q'}$ indicates that the partial derivatives are to be evaluated with t, q' fixed.

The propagator in (38.27) has been obtained for short times. It is, however, already more or less in its final form. We only have to evolve our short time approximation of the propagator according to (38.22)

$$K_{sc}(q'', q', t' + \delta t) = \sum_j \left| \det \frac{\partial q}{\partial q''} \right|_j^{1/2} e^{iR_j(q'', q, t')/\hbar - i\pi m_j(q'', q, t')/2} K(q, q', \delta t),$$

and we included here already the possibility that the phase becomes multi-valued, that is, that there is more than one path from q' to q'' . The topological index m_j =

Figure 38.4: Evolution of the semiclassical propagator. The configuration which corresponds to the initial conditions of the propagator is a Lagrangian manifold $q = q'$, that is, a plane parallel to the p axis. The hypothetical particles are thus initially all placed at q' but take on all possible momenta p' . The Jacobian matrix C (38.29) relates an initial volume element in momentum space dp' to a final configuration space volume dq .



$m_j(q'', q', t)$ is the number of singularities in the Jacobian along the trajectory j from q' to q'' . We can write $K_{sc}(q'', q', t' + \delta t)$ in closed form using the fact that $R(q'', q, t') + R(q, q', \delta t) = R(q'', q', t' + \delta t)$ and the multiplicativity of Jacobian determinants, that is

$$\det \left. \frac{\partial q}{\partial q''} \right|_t \det \left. \frac{\partial p'}{\partial q} \right|_{q', \delta t} = \det \left. \frac{\partial p'}{\partial q''} \right|_{q', t' + \delta t}. \quad (38.29)$$

The final form of the semiclassical or *Van Vleck propagator*, is thus

$$K_{sc}(q, q', t) = \sum_j \frac{1}{(2\pi i \hbar)^{D/2}} \left| \det \frac{\partial p'}{\partial q} \right|^{1/2} e^{iR_j(q, q', t)/\hbar - im_j \pi/2}. \quad (38.30)$$

This Van Vleck propagator is the essential ingredient of the semiclassical quantization to follow.

The apparent simplicity of the semiclassical propagator is deceptive. The wave function is not evolved simply by multiplying by a complex number of magnitude $\sqrt{\det \partial p' / \partial q}$ and phase $R(q, q', t)$; the more difficult task in general is to find the trajectories connecting q' and q in a given time t .

In addition, we have to treat the approximate propagator (38.30) with some care. Unlike the full quantum propagator, which satisfies the group property (36.11) exactly, the semiclassical propagator performs this only approximately, that is

$$K_{sc}(q, q', t_1 + t_2) \approx \int dq'' K_{sc}(q, q'', t_2) K_{sc}(q'', q', t_1). \quad (38.31)$$

The connection can be made explicit by the stationary phase approximation, sect. 37.2. Approximating the integral in (38.31) by integrating only over regions near points q'' at which the phase is stationary, leads to the stationary phase condition

$$\frac{\partial R(q, q'', t_2)}{\partial q''_i} + \frac{\partial R(q'', q', t_1)}{\partial q''_i} = 0. \quad (38.32)$$

Deterministic trajectories contribute whenever the final momentum for a path from q' to q'' and the initial momentum for a path from q'' to q coincide. Unlike the deterministic evolution of sect. 20.2, the semiclassical evolution is not an evolution by linear operator multiplication, but evolution supplemented by a stationary phase condition $p_{out} = p_{in}$ that matches up the deterministic momenta at each evolution step.

38.2.2 Free particle propagator

To develop some intuition about the above formalism, consider the case of a free particle. For a free particle the potential energy vanishes, the kinetic energy is $\frac{m}{2}\dot{q}^2$, and the Hamilton's principal function (38.12) is

$$R(q, q', t) = \frac{m(q - q')^2}{2t}. \quad (38.33)$$

The weight $\det \frac{\partial p'}{\partial q}$ from (38.28) can be evaluated explicitly, and the Van Vleck propagator is

$$K_{sc}(q, q', t) = \left(\frac{m}{2\pi i \hbar t} \right)^{D/2} e^{im(q-q')^2/2\hbar t}, \quad (38.34)$$

identical to the short time propagator (38.26), with $V(q) = 0$. This case is rather exceptional: for a free particle the semiclassical propagator turns out to be the exact quantum propagator $K(q, q', t)$, as can be checked by substitution in the Schrödinger equation (38.2). The Feynman path integral formalism uses this fact to construct an exact quantum propagator by integrating the free particle propagator (with $V(q)$ treated as constant for short times) along all possible (not necessarily deterministic) paths from q' to q .

remark 38.3

exercise 38.10

exercise 38.11

exercise 38.12

38.3 Semiclassical Green's function

This pathetic argument was pawned off to thousands of unsuspecting classes before the fundamental role of Green's functions was recognized; it is still to be found in several textbooks, and Professor Neanderthal loves it.

— *Ten Lessons*, by Gian-Carlo Rota [8]

So far we have derived semiclassical formulas for the time evolution of wave functions, that is, we obtained approximate solutions to the time dependent Schrödinger equation (36.1). Even though we assumed in the calculation a time-independent Hamiltonian of the special form (36.2), the derivation would lead to the same final result (38.30) were one to consider more complicated or explicitly time dependent Hamiltonians. The propagator is thus important when we are interested in finite time quantum mechanical effects. For time-independent Hamiltonians, the time dependence of the propagator as well as of wave functions is, however, essentially given in terms of the energy eigen-spectrum of the system, as in (36.8). It is therefore advantageous to switch from a time representation to an energy representation, that is from the propagator (36.10) to the energy dependent Green's function (36.14). A semiclassical approximation of the Green's function $G_{sc}(q, q', E)$ is given by the Laplace transform (36.14) of the Van Vleck propagator $K_{sc}(q, q', t)$:

$$G_{sc}(q, q', E) = \frac{1}{i\hbar} \int_0^\infty dt e^{iEt/\hbar} K_{sc}(q, q', t). \quad (38.35)$$

The expression as it stands is not very useful; in order to evaluate the integral, at least to the leading order in $\hbar$, we need to turn to the method of stationary phase again.

38.3.1 Stationary phase in higher dimensions

exercise 37.1

Generalizing the method of sect. 37.2 to d dimensions, consider stationary phase points fulfilling

$$\left. \frac{d}{dx_i} \Phi(x) \right|_{x=x_0} = 0 \quad \forall i = 1, \dots, d.$$

An expansion of the phase up to second order involves now the symmetric matrix of second derivatives of $\Phi(x)$, that is

$$D_{ij}(x_0) = \left. \frac{\partial^2}{\partial x_i \partial x_j} \Phi(x) \right|_{x=x_0}.$$

After choosing a suitable coordinate system which diagonalizes D , we can approximate the d -dimensional integral by d 1-dimensional Fresnel integrals; the stationary phase estimate of (A37.4) is then

$$I \approx \sum_{x_0} (2\pi i/s)^{d/2} |\det D(x_0)|^{-1/2} A(x_0) e^{is\Phi(x_0) - \frac{i\pi}{2} m(x_0)}, \quad (38.36)$$

where the sum runs over all stationary phase points x_0 of $\Phi(x)$ and $m(x_0)$ counts the number of negative eigenvalues of $D(x_0)$.

exercise 33.2

exercise 38.2

exercise 37.3

The stationary phase approximation is all that is needed for the semiclassical approximation, with the proviso that D in (38.36) has no zero eigenvalues.

38.3.2 Long trajectories

When evaluating the integral (38.35) approximately we have to distinguish between two types of contributions: those coming from stationary points of the phase and those coming from infinitesimally short times. The first type of contributions can be obtained by the stationary phase approximation and will be treated in this section. The latter originate from the singular behavior of the propagator for $t \rightarrow 0$ where the assumption that the amplitude changes slowly compared to the phase is not valid. The short time contributions therefore have to be treated separately, which we will do in sect. 38.3.3.

The stationary phase points t^* of the integrand in (38.35) are given by the condition

$$\frac{\partial}{\partial t} R(q, q', t^*) + E = 0. \quad (38.37)$$

We recognize this condition as the solution of (38.18), the time $t^* = t^*(q, q', E)$ in which a particle of energy E starting out in q' reaches q . Taking into account the second derivative of the phase evaluated at the stationary phase point,

$$R(q, q', t) + Et = R(q, q', t^*) + Et^* + \frac{1}{2}(t - t^*)^2 \frac{\partial^2}{\partial t^2} R(q, q', t^*) + \dots$$

the stationary phase approximation of the integral corresponding to a deterministic trajectory j in the Van Vleck propagator sum (38.30) yields

$$G_j(q, q', E) = \frac{1}{i\hbar(2i\pi\hbar)^{(D-1)/2}} \left| \det C_j \left(\frac{\partial^2 R_j}{\partial t^2} \right)^{-1} \right|^{1/2} e^{\frac{i}{\hbar} S_j - \frac{i\pi}{2} m_j}, \quad (38.38)$$

where $m_j = m_j(q, q', E)$ now includes a possible additional phase arising from the time stationary phase integration (37.16), and $C_j = C_j(q, q', t^*)$, $R_j = R_j(q, q', t^*)$ are evaluated at the transit time t^* . We re-express the phase in terms of the energy dependent action (38.16)

$$S(q, q', E) = R(q, q', t^*) + Et^*, \quad \text{with } t^* = t^*(q, q', E), \quad (38.39)$$

the Legendre transform of Hamilton's principal function. Note that the partial derivative of the action (38.39) with respect to q_i

$$\frac{\partial S(q, q', E)}{\partial q_i} = \frac{\partial R(q, q', t^*)}{\partial q_i} + \left(\frac{\partial R(q, q', t)}{\partial t^*} + E \right) \frac{\partial t}{\partial q_i}.$$

is equal to

$$\frac{\partial S(q, q', E)}{\partial q_i} = \frac{\partial R(q, q', t^*)}{\partial q_i}, \quad (38.40)$$

due to the stationary phase condition (38.37), so the definition of momentum as a partial derivative with respect to q remains unaltered by the Legendre transform from time to energy domain.

exercise 38.13

Next we will simplify the amplitude term in (38.38) and rewrite it as an explicit function of the energy. Consider the $[(D+1) \times (D+1)]$ matrix

$$D(q, q', E) = \begin{bmatrix} \frac{\partial^2 S}{\partial q' \partial q} & \frac{\partial^2 S}{\partial q' \partial E} \\ \frac{\partial^2 S}{\partial q \partial E} & \frac{\partial^2 S}{\partial E^2} \end{bmatrix} = \begin{bmatrix} -\frac{\partial p'}{\partial q} & -\frac{\partial p'}{\partial E} \\ \frac{\partial t}{\partial q} & \frac{\partial t}{\partial E} \end{bmatrix}, \quad (38.41)$$

where $S = S(q, q', E)$ and we used (38.14–38.17) here to obtain the left hand side of (38.41). The minus signs follow from the definition of (38.15), which implies that $S(q, q', E) = -S(q', q, E)$. Note that D is nothing but the Jacobian matrix of the coordinate transformation $(q, E) \rightarrow (p', t)$ for fixed q' . We can therefore use the multiplication rules of determinants of Jacobians, which are just ratios of volume elements, to obtain

$$\begin{aligned} \det D &= (-1)^{D+1} \left(\det \frac{\partial(p', t)}{\partial(q, E)} \right)_{q'} = (-1)^{D+1} \left(\det \frac{\partial(p', t)}{\partial(q, t)} \frac{\partial(q, t)}{\partial(q, E)} \right)_{q'} \\ &= (-1)^{D+1} \left(\det \frac{\partial p'}{\partial q} \right)_{t, q'} \left(\det \frac{\partial t}{\partial E} \right)_{q', q} = \det C \left(\frac{\partial^2 R}{\partial t^2} \right)^{-1}. \end{aligned}$$

We use here the notation $(\det \cdot)_{q', t}$ for a Jacobian determinant with partial derivatives evaluated at t, q' fixed, and likewise for other subscripts. Using the relation (38.19) which relates the term $\frac{\partial t}{\partial E}$ to $\partial_t^2 R$ we can write the determinant of D as

a product of the Van Vleck determinant (38.28) and the amplitude factor arising from the stationary phase approximation. The amplitude in (38.38) can thus be interpreted as the determinant of a Jacobian of a coordinate transformation which includes time and energy as independent coordinates. This causes the increase in the dimensionality of the matrix D relative to the Van Vleck determinant (38.28).

We can now write down the semiclassical approximation of the contribution of the j th trajectory to the Green's function (38.38) in explicitly energy dependent form:

$$G_j(q, q', E) = \frac{1}{i\hbar(2i\pi\hbar)^{(D-1)/2}} |\det D_j|^{1/2} e^{iS_j - \frac{i\pi}{2}m_j}. \quad (38.42)$$

However, this is still not the most convenient form of the Green's function.

The trajectory contributing to $G_j(q, q', E)$ is constrained to a given energy E , and will therefore be on a phase space manifold of constant energy, that is $H(q, p) = E$. Writing this condition as a partial differential equation for $S(q, q', E)$, that is

$$H(q, \frac{\partial S}{\partial q}) = E,$$

one obtains

$$\begin{aligned} \frac{\partial}{\partial q'_i} H(q, p) &= 0 = \frac{\partial H}{\partial p_j} \frac{\partial p_j}{\partial q'_i} = \dot{q}_j \frac{\partial^2 S}{\partial q_j \partial q'_i} \\ \frac{\partial}{\partial q_i} H(q', p') &= 0 = \frac{\partial^2 S}{\partial q_i \partial q'_j} \dot{q}'_j, \end{aligned} \quad (38.43)$$

that is the sub-matrix $\partial^2 S / \partial q_i \partial q'_j$ has (left- and right-) eigenvectors corresponding to an eigenvalue 0. Rotate the local coordinate system at the either end of the trajectory

$$(q_1, q_2, q_3, \dots, q_d) \rightarrow (q_{\parallel}, q_{\perp 1}, q_{\perp 2}, \dots, q_{\perp(D-1)})$$

so that one axis points along the trajectory and all others are perpendicular to it

$$(\dot{q}_1, \dot{q}_2, \dot{q}_3, \dots, \dot{q}_d) \rightarrow (\dot{q}, 0, 0, \dots, 0).$$

With such local coordinate systems at both ends, with the longitudinal coordinate axis $q_{\parallel}$ pointing along the velocity vector of magnitude $\dot{q}$, the stability matrix of $S(q, q', E)$ has a column and a row of zeros as (38.43) takes form

$$\dot{q} \frac{\partial^2 S}{\partial q_{\parallel} \partial q'_i} = \frac{\partial^2 S}{\partial q_i \partial q'_{\parallel}} \dot{q}' = 0.$$

The initial and final velocities are non-vanishing except for points $|\dot{q}| = 0$. These are the turning points (where all energy is potential), and we assume that neither q nor q' is a turning point (in our application - periodic orbits - we can always chose $q = q'$ not a turning point). In the local coordinate system with one axis along

the trajectory and all other perpendicular to it, the determinant of (38.41) is of the form

$$\det D(q, q', E) = (-1)^{D+1} \det \begin{pmatrix} 0 & 0 & \frac{\partial^2 S}{\partial E \partial q'_\parallel} \\ 0 & \frac{\partial^2 S}{\partial q_\perp \partial q'_\perp} & * \\ \frac{\partial^2 S}{\partial q_\parallel \partial E} & * & * \end{pmatrix}. \quad (38.44)$$

The corner entries can be evaluated using (38.17)

$$\frac{\partial^2 S}{\partial q_\parallel \partial E} = \frac{\partial}{\partial q_\parallel} t = \frac{1}{\dot{q}}, \quad \frac{\partial^2 S}{\partial E \partial q'_\parallel} = \frac{1}{\dot{q}'}.$$

As the $q_\parallel$ axis points along the velocity direction, velocities $\dot{q}, \dot{q}'$ are by construction almost always positive non-vanishing numbers. In this way the determinant of the $[(D+1) \times (D+1)]$ dimensional matrix $D(q, q', E)$ can be reduced to the determinant of a $[(D-1) \times (D-1)]$ dimensional *transverse* matrix $D_\perp(q, q', E)$

$$\begin{aligned} \det D(q, q', E) &= \frac{1}{\dot{q}\dot{q}'} \det D_\perp(q, q', E) \\ D_\perp(q, q', E)_{ik} &= -\frac{\partial^2 S(q, q', E)}{\partial q_{\perp i} \partial q'_{\perp k}}. \end{aligned} \quad (38.45)$$

Putting everything together we obtain the j th trajectory contribution to the semi-classical Green's function

$$G_j(q, q', E) = \frac{1}{i\hbar(2\pi i\hbar)^{(D-1)/2}} \frac{1}{|\dot{q}\dot{q}'|^{1/2}} \left| \det D_\perp^j \right|^{1/2} e^{\frac{i}{\hbar} S_j - \frac{i\pi}{2} m_j}, \quad (38.46)$$

exercise 38.15

where the topological index $m_j = m_j(q, q', E)$ now counts the number of changes of sign of $\det D_\perp^j$ along the trajectory j which connects q' to q at energy E .

The endpoint velocities $\dot{q}, \dot{q}'$ also depend on (q, q', E) and the trajectory j .

38.3.3 Short trajectories

The stationary phase method cannot be used when t^* is small, both because we cannot extend the integration in (37.16) to $-\infty$, and because the amplitude of $K(q, q', t)$ is divergent. In this case we have to evaluate the integral involving the short time form of the exact quantum mechanical propagator (38.26)

$$G_0(q, q', E) = \frac{1}{i\hbar} \int_0^\infty dt \left(\frac{m}{2\pi i\hbar t} \right)^{D/2} e^{\frac{i}{\hbar} \left(\frac{m(q-q')^2}{2t} - V(q)t + Et \right)}. \quad (38.47)$$

By introducing a dimensionless variable $\tau = t \sqrt{2m(E - V(q))}/m|q - q'|$, the integral can be rewritten as

$$G_0(q, q', E) = \frac{m}{i\hbar^2(2\pi i)^{D/2}} \left(\frac{\sqrt{2m(E - V)}}{\hbar|q - q'|} \right)^{\frac{D}{2}-1} \int_0^\infty \frac{d\tau}{\tau^{D/2}} e^{\frac{i}{2\hbar} S_0(q, q', E)(\tau+1/\tau)},$$

where $S_0(q, q', E) = \sqrt{2m(E - V)}|q - q'|$ is the short distance form of the action. Using the integral representation of the Hankel function of first kind

$$H_\nu^+(z) = -\frac{i}{\pi} e^{-i\nu\pi/2} \int_0^\infty e^{\frac{1}{2}iz(\tau+1/\tau)} \tau^{-\nu-1} d\tau$$

we can write the short distance form of the Green's function as

$$G_0(q, q', E) \approx -\frac{im}{2\hbar^2} \left(\frac{\sqrt{2m(E - V)}}{2\pi\hbar|q - q'|} \right)^{\frac{D-2}{2}} H_{\frac{D-2}{2}}^+(S_0(q, q', E)/\hbar). \quad (38.48)$$

Hankel functions are standard, and their short wavelength asymptotics is described in standard reference books. The short distance Green's function approximation is valid when $S_0(q, q', E) \leq \hbar$.

Résumé

The aim of the semiclassical or short-wavelength methods is to approximate a solution of the Schrödinger equation with a semiclassical wave function

$$\psi_{sc}(q, t) = \sum_j A_j(q, t) e^{iR_j(q, t)/\hbar},$$

accurate to the leading order in $\hbar$. Here the sum is over all deterministic trajectories that connect the initial point q' to the final point q in time t . “Semi-” refers to $\hbar$, the quantum unit of phase in the exponent. The quantum mechanics enters only through this atomic scale, in units of which the variation of the phase across the potential is assumed to be large. “-classical” refers to the rest - both the amplitudes $A_j(q, t)$ and the phases $R_j(q, t)$ - which are determined by the classical Hamilton-Jacobi equations.

In the semiclassical approximation the quantum time evolution operator is given by the *semiclassical propagator*

$$K_{sc}(q, q', t) = \frac{1}{(2\pi i\hbar)^{D/2}} \sum_j \left| \det \frac{\partial p'}{\partial q} \right|_j^{1/2} e^{\frac{i}{\hbar} R_j - \frac{i\pi}{2} m_j},$$

where the topological index $m_j(q, q', t)$ counts the number of the direction reversal along the j th deterministic trajectory that connects $q' \rightarrow q$ in time t . Until very recently it was not possible to resolve quantum evolution on quantum time scales (such as one revolution of electron around a nucleus) - physical measurements are almost always done at time scales asymptotically large compared to the intrinsic quantum time scale. Formally this information is extracted by means of a Laplace transform of the propagator which yields the energy dependent *semiclassical Green's function*

$$\begin{aligned} G_{sc}(q, q', E) &= G_0(q, q', E) + \sum_j G_j(q, q', E) \\ G_j(q, q', E) &= \frac{1}{i\hbar(2\pi i\hbar)^{\frac{(D-1)}{2}}} \left| \frac{1}{\dot{q}\dot{q}'} \det \frac{\partial p'_\perp}{\partial q_\perp} \right|_j^{1/2} e^{\frac{i}{\hbar} S_j - \frac{i\pi}{2} m_j} \end{aligned} \quad (38.49)$$

where $G_0(q, q', E)$ is the contribution of short trajectories with $S_0(q, q', E) \leq \hbar$, while the sum is over the contributions of long trajectories (38.46) going from q' to q with fixed energy E , with $S_j(q, q', E) \gg \hbar$.

Commentary

Remark 38.1. Limit $\hbar \rightarrow 0$. The semiclassical limit “ $\hbar \rightarrow 0$ ” discussed in sect. 38.1 is a shorthand notation for the limit in which typical quantities like the actions R or S in semiclassical expressions for the propagator or the Green’s function become large compared to $\hbar$. In the world that we live in the quantity $\hbar$ is a fixed physical constant whose value is $1.054571800 \cdot 10^{-34}$ Js.

Remark 38.2. Madelung’s fluid dynamics. Already Schrödinger [9–12] noted that

$$\rho = \rho(q, t) := A^2 = \psi^* \psi$$

plays the role of a density, and that the gradient of R may be interpreted as a local semiclassical momentum, as the momentum density is

$$\psi(q, t)^* (-i\hbar \frac{\partial}{\partial q}) \psi(q, t) = -i\hbar A \frac{\partial A}{\partial q} + \rho \frac{\partial R}{\partial q}.$$

A very different interpretation of (38.3–38.4) has been given by Madelung [6], and then built upon by Bohm [2] and others [5, 9–12]. Keeping the $\hbar$ dependent term in (38.3), the ordinary differential equations driving the flow (38.10) have to be altered; if the Hamiltonian can be written as kinetic plus potential term $V(q)$ as in (36.2), the $\hbar^2$ term modifies the p equation of motion as

$$\dot{p}_i = -\frac{\partial}{\partial q_i} (V(q) + Q(q, t)), \quad (38.50)$$

where, for the example at hand,

$$Q(q, t) = -\frac{\hbar^2}{2m} \frac{1}{\sqrt{\rho}} \frac{\partial^2}{\partial q^2} \sqrt{\rho} \quad (38.51)$$

interpreted by Bohm [2] as the “quantum potential.” Madelung observed that Hamilton’s equation for the momentum (38.50) can be rewritten as

$$\frac{\partial v_i}{\partial t} + \left(v \cdot \frac{\partial}{\partial q} \right) v_i = -\frac{1}{m} \frac{\partial V}{\partial q_i} - \frac{1}{m\rho} \frac{\partial}{\partial q_j} \sigma_{ij}, \quad (38.52)$$

where $\sigma_{ij} = \frac{\hbar^2 \rho}{4m} \frac{\partial^2 \ln \rho}{\partial q_i \partial q_j}$ is the ‘pressure’ stress tensor, $v_i = p_i/m$, and $\rho = A^2$ as defined [9–12] in sect. 38.1.3. We recall that the Eulerian $\frac{\partial}{\partial t} + \frac{\partial q_i}{\partial t} \frac{\partial}{\partial q_i}$ is the ordinary derivative of Lagrangian mechanics, that is $\frac{d}{dt}$. For comparison, the Euler equation for classical hydrodynamics is

$$\frac{\partial v_i}{\partial t} + \left(v \cdot \frac{\partial}{\partial q} \right) v_i = -\frac{1}{m} \frac{\partial V}{\partial q_i} - \frac{1}{m\rho} \frac{\partial}{\partial q_j} (p\delta_{ij}),$$

where $p\delta_{ij}$ is the pressure tensor.

The deterministic dynamics corresponding to quantum evolution is thus that of an “hypothetical fluid” experiencing $\hbar$ and ρ dependent stresses. The “hydrodynamic” interpretation of quantum mechanics has, however, not been very fruitful in practice.

Remark 38.3. Path integrals. The semiclassical propagator (38.30) can also be derived from Feynman’s path integral formalism. Dirac was the first to discover that in the short-time limit the quantum propagator (38.34) is exact. Feynman noted in 1946 that one can construct the exact propagator of the quantum Schrödinger equation by formally summing over all possible (and emphatically not deterministic) paths from q' to q . Gutzwiller started from the path integral to rederive Van Vleck’s semiclassical expression for the propagator; Van Vleck’s original derivation is very much in the spirit of what has presented in this chapter. He did, however, not consider the possibility of the formation of caustics or folds of Lagrangian manifolds and thus did not include the topological phases in his semiclassical expression for the propagator. Some 40 years later Gutzwiller [3] added the topological indices when deriving the semiclassical propagator from Feynman’s path integral by stationary phase conditions.

Remark 38.4. Applications of the semiclassical Green’s function. The semiclassical Green’s function is the starting point of the semiclassical approximation in many applications. The generic semiclassical strategy is to express physical quantities (for example scattering amplitudes and cross section in scattering theory, oscillator strength in spectroscopy, and conductance in mesoscopic physics) in terms of the exact Green’s function and then replace it with the semiclassical formula.

Remark 38.5. The quasiclassical approximation The *quasiclassical* approximation was introduced by Maslov [7]. The term ‘quasiclassical’ is more appropriate than semiclassical since the Maslov type description leads to a pure deterministic evolution operator in a natural way. Following mostly ref. [1], we give a summary of the quasiclassical approximation, which was worked out by Maslov [7] in this form. One additional advantage of this description is that the wave function evolves along one single deterministic trajectory and we do not have to compute sums over increasing numbers of deterministic trajectories as in computations involving Van Vleck formula [4].

References

- [1] V. I. Arnol’d, “Characteristic class entering in quantization conditions”, *Funct. Anal. Appl.* **1**, 1–13 (1967).
- [2] D. Bohm, “A suggested interpretation of the quantum theory in terms of “hidden” variables. I”, *Phys. Rev.* **85**, 166–179 (1952).
- [3] M. C. Gutzwiller, “Phase-integral approximation in momentum space and the bound states of an atom”, *J. Math. Phys.* **8**, 1979–2000 (1967).
- [4] E. J. Heller, S. Tomsovic, and M. A. Sepúlveda, “Time domain approach to semiclassical dynamics: Breaking the log time barrier”, *Chaos*, 105–115 (1992).
- [5] P. R. Holland, *The Quantum Theory of Motion: An Account of the de Broglie-Bohm Causal Interpretation of Quantum Mechanics* (Cambridge Univ. Press, Cambridge UK, 1993).
- [6] E. Madelung, “Quantentheorie in hydrodynamischer Form”, *Z. f. Physik* **40**, 322–326 (1927).

- [7] V. P. Maslov and M. V. Fedoriuk, *Semi-Classical Approximation in Quantum Mechanics* (Reidel, Boston, 1981).
- [8] G.-C. Rota, *Ten lessons I wish I had learned before I started teaching differential equations*, tech. rep., Simmons College Meeting of Math. Assoc. of America (Massachusetts Inst. of Technology, 1997).
- [9] E. Schrödinger, “*Quantisierung als Eigenwertproblem*”, *Ann. Phys.* **384**, 361–376 (1926).
- [10] E. Schrödinger, “*Quantisierung als Eigenwertproblem*”, *Ann. Phys.* **384**, 489–527 (1926).
- [11] E. Schrödinger, “*Quantisierung als Eigenwertproblem*”, *Ann. Phys.* **385**, 437–490 (1926).
- [12] E. Schrödinger, “*Quantisierung als Eigenwertproblem*”, *Ann. Phys.* **386**, 109–139 (1926).

38.4 Exercises

- 38.1. **Dirac delta function, Gaussian representation.** Consider the Gaussian distribution function

$$\delta_\sigma(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2\sigma^2}.$$

Show that in $\sigma \rightarrow 0$ limit this is the Dirac delta function

$$\int_{\mathcal{M}} dx \delta(x) = 1 \text{ if } 0 \in \mathcal{M}, \text{ zero otherwise.}$$

- 38.2. **Stationary phase approximation in higher dimensions.** All semiclassical approximations are based on saddle point evaluations of integrals of type

$$I = \int d^D x A(x) e^{i\Phi(x)/\hbar} \quad (38.53)$$

for small values of $\hbar$. Obtain the stationary phase estimate

$$I \approx \sum_n A(x_n) e^{i\Phi(x_n)/\hbar} \frac{(2\pi i \hbar)^{D/2}}{\sqrt{\det \mathbf{D}^2 \Phi(x_n)}},$$

where $\mathbf{D}^2 \Phi(x_n)$ denotes the second derivative matrix.

- 38.3. **Schrödinger equation in the Madelung form.** Verify the decomposition of Schrödinger equation into real and imaginary parts, eqs. (38.3) and (38.4).

- 38.4. **Transport equations.**  Write the wave-function in the asymptotic form

$$\psi(q, t) = e^{\frac{i}{\hbar} R(x, t) + \frac{i}{\hbar} \epsilon t} \sum_{n \geq 0} (i\hbar)^n A_n(x, t).$$

Derive the transport equations for the A_n by substituting this into the Schrödinger equation and then collecting terms by orders of $\hbar$. Note that equation for $\dot{A}_n$ only requires knowledge of A_{n-1} and R .

- 38.5. **Easy examples of the Hamilton's principal function.** Calculate $R(q, q', t)$ for

- a) a D -dimensional free particle
- b) a 3-dimensional particle in constant magnetic field
- c) a 1-dimensional harmonic oscillator.

(continuation: exercise 38.13.)

- 38.6. **1-dimensional harmonic oscillator.** Take a 1-dimensional harmonic oscillator $U(q) = \frac{1}{2} k q^2$. Take a WKB wave function of form $A(q, t) = a(t)$ and $R(q, t) = r(t) + b(t)q + c(t)q^2$, where $r(t)$, $a(t)$, $b(t)$ and $c(t)$ are time dependent coefficients. Derive ordinary differential equations by using (38.3) and (38.4) and solve them. (continuation: exercise 38.9.)

- 38.7. **1-dimensional linear potential.** Take a 1-dimensional linear potential $U(q) = -Fq$. Take a WKB wave function of form $A(q, t) = a(t)$ and $R(q, t) = r(t) + b(t)q + c(t)q^2$, where $r(t)$, $a(t)$, $b(t)$ and $c(t)$ are time dependent coefficients. Derive and solve the ordinary differential equations from (38.3) and (38.4).

- 38.8. ***D*-dimensional quadratic potentials.** Generalize the above method to general *D*-dimensional quadratic potentials.
- 38.9. **Time evolution of *R*.** (continuation of exercise 38.6) Calculate the time evolution of $R(q, 0) = a + bq + cq^2$ for a 1-dimensional harmonic oscillator using (38.12) and (38.14).
- 38.10. ***D*-dimensional free particle propagator.** Verify the results in sect. 38.2.2; show explicitly that (38.34), the semiclassical Van Vleck propagator in *D* dimensions, solves the Schrödinger's equation.
- 38.11. **Propagator, charged particle in constant magnetic field.** Calculate the semiclassical propagator for a charged particle in constant magnetic field in 3 dimensions. Verify that the semiclassical expression coincides with the exact solution.
- 38.12. **1-dimensional harmonic oscillator propagator.** Calculate the semiclassical propagator for a 1-dimensional harmonic oscillator and verify that it is identical to the exact quantum propagator.
- 38.13. **Free particle action.** Calculate the energy dependent action for a free particle, a charged particle in a constant magnetic field and for the harmonic oscillator.
- 38.14. **Zero length orbits.**  Derive the deterministic trace (21.1) rigorously and either add the $t \rightarrow 0_+$ zero length contribution to the trace formula, or show that it vanishes. Send us a reprint of *Phys. Rev. Lett.* with the correct derivation.
- 38.15. **Free particle semiclassical Green's functions.** Calculate the semiclassical Green's functions for the systems of exercise 38.13.

Chapter 39

Semiclassical quantization

(G. Vattay, G. Tanner and P. Cvitanović)

WE DERIVE HERE the Gutzwiller trace formula and the semiclassical zeta function, the central results of the semiclassical quantization of deterministically chaotic systems. In chapter 40 we will rederive these formulas for the case of scattering in open systems. Quintessential wave mechanics effects such as creeping, diffraction and tunneling will be taken up in chapter 43.

39.1 Trace formula

Our next task is to evaluate the Green's function trace (36.15) in the semiclassical approximation. The trace

$$\mathrm{Tr} G_{sc}(E) = \int d^D q G_{sc}(q, q, E) = \mathrm{Tr} G_0(E) + \sum_j \int d^D q G_j(q, q, E)$$

receives contributions from “long” deterministic trajectories labeled by j which start and end in q after finite time, and the “zero length” trajectories whose lengths approach zero as $q' \rightarrow q$.

First, we work out the contributions coming from the finite time *returning* deterministic orbits, i.e., trajectories that originate and end at a given configuration point q . As we are identifying q with q' , taking of a trace involves (still another!) stationary phase condition in the $q' \rightarrow q$ limit,

$$\left. \frac{\partial S_j(q, q', E)}{\partial q_i} \right|_{q'=q} + \left. \frac{\partial S_j(q, q', E)}{\partial q'_i} \right|_{q'=q} = 0,$$

meaning that the initial and final momenta (38.40) of contributing trajectories should coincide

$$p_i(q, q, E) - p'_i(q, q, E) = 0, \quad q \in j\text{th periodic orbit}, \quad (39.1)$$

Figure 39.1: A returning trajectory in the configuration space. The orbit is periodic in the full phase space only if the initial and the final momenta of a returning trajectory coincide as well.



Figure 39.2: A romanticized sketch of $S_p(E) = S(q, q, E) = \oint p(q, E) dq$ landscape orbit. Unstable periodic orbits traverse isolated ridges and saddles of the mountainous landscape of the action $S(q_{||}, q_{\perp}, E)$. Along a periodic orbit $S_p(E)$ is constant; in the transverse directions it generically changes quadratically.



so the trace receives contributions only from those long deterministic trajectories which are *periodic* in the full phase space.

For a periodic orbit the natural coordinate system is the intrinsic one, with $q_{||}$ axis pointing in the $\dot{q}$ direction along the orbit, and $q_{\perp}$, the rest of the coordinates transverse to $\dot{q}$. The j th periodic orbit contribution to the trace of the semiclassical Green's function in the intrinsic coordinates is

$$\text{Tr } G_j(E) = \frac{1}{i\hbar(2\pi\hbar)^{(d-1)/2}} \oint_j \frac{dq_{||}}{\dot{q}} \int_j d^{d-1} q_{\perp} |\det D_{\perp}^j|^{1/2} e^{\frac{i}{\hbar} S_j - \frac{i\pi}{2} m_j},$$

where the integration in $q_{||}$ goes from 0 to L_j , the geometric length of small tube around the orbit in the configuration space. As always, in the stationary phase approximation we worry only about the fast variations in the phase $S_j(q_{||}, q_{\perp}, E)$, and assume that the density varies smoothly and is well approximated by its value $D_{\perp}^j(q_{||}, 0, E)$ on the deterministic trajectory, $q_{\perp} = 0$. The topological index $m_j(q_{||}, q_{\perp}, E)$ is an integer which does not depend on the initial point $q_{||}$ and not change in the infinitesimal neighborhood of an isolated periodic orbit, so we set $m_j(E) = m_j(q_{||}, q_{\perp}, E)$.

The transverse integration is again carried out by the stationary phase method, with the phase stationary on the periodic orbit, $q_{\perp} = 0$. The result of the transverse integration can depend only on the parallel coordinate

$$\text{Tr } G_j(E) = \frac{1}{i\hbar} \oint \frac{dq_{||}}{\dot{q}} \left| \frac{\det D_{\perp j}(q_{||}, 0, E)}{\det D'_{\perp j}(q_{||}, 0, E)} \right|^{1/2} e^{\frac{i}{\hbar} S_j - \frac{i\pi}{2} m_j},$$

where the new determinant in the denominator, $\det D'_{\perp j} =$

$$\det \left(\frac{\partial^2 S(q, q', E)}{\partial q_{\perp i} \partial q_{\perp j}} + \frac{\partial^2 S(q, q', E)}{\partial q'_{\perp i} \partial q_{\perp j}} + \frac{\partial^2 S(q, q', E)}{\partial q_{\perp i} \partial q'_{\perp j}} + \frac{\partial^2 S(q, q', E)}{\partial q'_{\perp i} \partial q'_{\perp j}} \right),$$

is the determinant of the second derivative matrix coming from the stationary phase integral in transverse directions.

The ratio $\det D_{\perp j} / \det D'_{\perp j}$ is here to enforce the periodic boundary condition for the semiclassical Green's function evaluated on a periodic orbit. It can be given a meaning in terms of the monodromy matrix of the periodic orbit by following observations

$$\begin{aligned} \det D_{\perp} &= \left\| \frac{\partial p'_{\perp}}{\partial q_{\perp}} \right\| = \left\| \frac{\partial(q'_{\perp}, p'_{\perp})}{\partial(q_{\perp}, q'_{\perp})} \right\| \\ \det D'_{\perp} &= \left\| \frac{\partial p_{\perp}}{\partial q_{\perp}} - \frac{\partial p'_{\perp}}{\partial q_{\perp}} + \frac{\partial p_{\perp}}{\partial q'_{\perp}} - \frac{\partial p'_{\perp}}{\partial q'_{\perp}} \right\| = \left\| \frac{\partial(q_{\perp} - q'_{\perp}, p_{\perp} - p'_{\perp})}{\partial(q_{\perp}, q'_{\perp})} \right\|. \end{aligned}$$

Defining the $2(D - 1)$ -dimensional transverse vector $x_{\perp} = (q_{\perp}, p_{\perp})$ in the full phase space we can express the ratio

$$\begin{aligned} \frac{\det D'_{\perp}}{\det D_{\perp}} &= \left\| \frac{\partial(q_{\perp} - q'_{\perp}, p_{\perp} - p'_{\perp})}{\partial(q'_{\perp}, p'_{\perp})} \right\| = \left\| \frac{\partial(x_{\perp} - x'_{\perp})}{\partial x'_{\perp}} \right\| \\ &= \det(M - \mathbf{1}), \end{aligned} \quad (39.2)$$

in terms of the monodromy matrix M for a surface of section transverse to the orbit within the constant energy $E = H(q, p)$ shell.

The deterministic periodic orbit action $S_j(E) = \oint p(q_{\parallel}, E) dq_{\parallel}$ is an integral around a loop defined by the periodic orbit, and does not depend on the starting point $q_{\parallel}$ along the orbit, see figure 39.2. The eigenvalues of the monodromy matrix are also independent of where M_j is evaluated along the orbit, so $\det(1 - M_j)$ can also be taken out of the $q_{\parallel}$ integral

$$\text{Tr } G_j(E) = \frac{1}{i\hbar} \sum_j \frac{1}{|\det(1 - M_j)|^{1/2}} e^{r(\frac{i}{\hbar} S_j - \frac{i\pi}{2} m_j)} \oint \frac{dq_{\parallel}}{\dot{q}_{\parallel}}.$$

Here we have assumed that M_j has no marginal eigenvalues. The determinant of the monodromy matrix, the action $S_p(E) = \oint p(q_{\parallel}, E) dq_{\parallel}$ and the topological index are all deterministic invariants of the periodic orbit. The integral in the parallel direction we now do exactly.

First, we take into account the fact that any repeat of a periodic orbit is also a periodic orbit. The action and the topological index are additive along the trajectory, so for r th repeat they simply get multiplied by r . The monodromy matrix of the r th repeat of a prime cycle p is (by the chain rule for derivatives) M_p^r , where M_p is the prime cycle monodromy matrix. Let us denote the time period of the prime cycle p , the single, shortest traversal of a periodic orbit by T_p . The remaining integral can be carried out by change of variables $dt = dq_{\parallel} / \dot{q}(t)$

$$\int_0^{L_p} \frac{dq_{\parallel}}{\dot{q}(t)} = \int_0^{T_p} dt = T_p.$$

Note that the spatial integral corresponds to a *single* traversal. If you do not see why this is so, rethink the derivation of the deterministic trace formula (21.19) -

that derivation takes only three pages of text. Regrettably, in the quantum case we do not know of an honest derivation that takes less than 30 pages. The final result, the *Gutzwiller trace formula*

$$\mathrm{Tr} G_{sc}(E) = \mathrm{Tr} G_0(E) + \frac{1}{i\hbar} \sum_p T_p \sum_{r=1}^{\infty} \frac{1}{|\det(1 - M_p^r)|^{1/2}} e^{r(\frac{i}{\hbar} S_p - \frac{ir}{2} m_p)}, \quad (39.3)$$

an expression for the trace of the semiclassical Green's function in terms of periodic orbits, is beautiful in its simplicity and elegance.

The topological index $m_p(E)$ counts the number of changes of sign of the matrix of second derivatives evaluated along the prime periodic orbit p . By now we have gone through so many stationary phase approximations that you have surely lost track of what the total $m_p(E)$ actually is. The rule is this: The topological index of a closed curve in a $2D$ phase space is the sum of the number of times the partial derivatives $\frac{\partial p_i}{\partial q_i}$ for each dual pair (q_i, p_i) , $i = 1, 2, \dots, D$ (no sum on i) change their signs as one goes once around the curve.

39.1.1 Average density of states

We still have to evaluate $\mathrm{Tr} G_0(E)$, the contribution coming from the infinitesimal trajectories. The real part of $\mathrm{tr} G_0(E)$ is infinite in the $q' \rightarrow q$ limit, so it makes no sense to write it down explicitly here. However, the imaginary part is finite, and plays an important role in the density of states formula, which we derive next.

The semiclassical contribution to the density of states (36.15) is given by the imaginary part of the Gutzwiller trace formula (39.3) multiplied with $-1/\pi$. The contribution coming from the zero length trajectories is the imaginary part of (38.48) for $q' \rightarrow q$ integrated over the configuration space

$$d_0(E) = -\frac{1}{\pi} \int d^D q \mathrm{Im} G_0(q, q, E),$$

The resulting formula has a pretty interpretation; it estimates the number of quantum states that can be accommodated up to the energy E by counting the available quantum cells in the phase space. This number is given by the *Weyl rule*, as the ratio of the phase space volume bounded by energy E divided by h^D , the volume of a quantum cell,

$$N_{sc}(E) = \frac{1}{h^D} \int d^D p d^D q \Theta(E - H(q, p)). \quad (39.4)$$

where $\Theta(x)$ is the Heaviside function (36.20). $N_{sc}(E)$ is an estimate of the spectral staircase (36.19), so its derivative yields the average density of states

$$d_0(E) = \frac{d}{dE} N_{sc}(E) = \frac{1}{h^D} \int d^D p d^D q \delta(E - H(q, p)), \quad (39.5)$$

precisely the semiclassical result (39.6). For Hamiltonians of type $p^2/2m + V(q)$, the energy shell volume in (39.5) is a sphere of radius $\sqrt{2m(E - V(q))}$. The

surface of a d -dimensional sphere of radius r is $\pi^{d/2} r^{d-1} / \Gamma(d/2)$, so the average density of states is given by

exercise 39.2

$$d_0(E) = \frac{2m}{\hbar^D 2^d \pi^{D/2} \Gamma(D/2)} \int_{V(q) < E} d^D q [2m(E - V(q))]^{D/2-1}, \quad (39.6)$$

and

$$N_{sc}(E) = \frac{1}{h^D} \frac{\pi^{D/2}}{\Gamma(1 + D/2)} \int_{V(q) < E} d^D q [2m(E - V(q))]^{D/2}. \quad (39.7)$$

Physically this means that at a fixed energy the phase space can support $N_{sc}(E)$ distinct eigenfunctions; anything finer than the quantum cell h^D cannot be resolved, so the quantum phase space is effectively finite dimensional. The average density of states is of a particularly simple form in one spatial dimension

exercise 39.3

$$d_0(E) = \frac{T(E)}{2\pi\hbar}, \quad (39.8)$$

where $T(E)$ is the period of the periodic orbit of fixed energy E . In two spatial dimensions the average density of states is

$$d_0(E) = \frac{m\mathcal{A}(E)}{2\pi\hbar^2}, \quad (39.9)$$

where $\mathcal{A}(E)$ is the deterministically allowed area of configuration space for which $V(q) < E$.

exercise 39.4

The semiclassical density of states is a sum of the average density of states and the oscillation of the density of states around the average, $d_{sc}(E) = d_0(E) + d_{osc}(E)$, where

$$d_{osc}(E) = \frac{1}{\pi\hbar} \sum_p T_p \sum_{r=1}^{\infty} \frac{\cos(rS_p(E)/\hbar - rm_p\pi/2)}{|\det(1 - M_p^r)|^{1/2}} \quad (39.10)$$

follows from the trace formula (39.3).

39.1.2 Regularization of the trace

The real part of the $q' \rightarrow q$ zero length Green's function (38.48) is ultraviolet divergent in dimensions $d > 1$, and so is its formal trace (36.15). The short distance behavior of the real part of the Green's function can be extracted from the real part of (38.48) by using the Bessel function expansion for small z

$$Y_\nu(z) \approx \begin{cases} -\frac{1}{\pi} \Gamma(\nu) \left(\frac{z}{2}\right)^{-\nu} & \text{for } \nu \neq 0 \\ \frac{2}{\pi} (\ln(z/2) + \gamma) & \text{for } \nu = 0 \end{cases},$$

where $\gamma = 0.577\dots$ is the Euler constant. The real part of the Green's function for short distance is dominated by the singular part

$$G_{sing}(|q - q'|, E) = \begin{cases} -\frac{m}{2\hbar^2 \pi^{d/2}} \Gamma((d-2)/2) \frac{1}{|q - q'|^{d-2}} & \text{for } d \neq 2 \\ \frac{m}{2\pi\hbar^2} (\ln(2m(E - V)|q - q'|/2\hbar) + \gamma) & \text{for } d = 2 \end{cases}.$$

The *regularized* Green's function

$$G_{reg}(q, q', E) = G(q, q', E) - G_{sing}(|q - q'|, E)$$

is obtained by subtracting the $q' \rightarrow q$ ultraviolet divergence. For the regularized Green's function the Gutzwiller trace formula is

$$\text{Tr } G_{reg}(E) = -i\pi d_0(E) + \frac{1}{i\hbar} \sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r(\frac{i}{\hbar} S_p(E) - \frac{i\pi}{2} m_p(E))}}{|\det(1 - M_p^r)|^{1/2}}. \quad (39.11)$$

Now you stand where Gutzwiller stood in 1990. You hold the trace formula in your hands. You have no clue how good is the $\hbar \rightarrow 0$ approximation, how to take care of the sum over an infinity of periodic orbits, and whether the formula converges at all.

39.2 Semiclassical spectral determinant

The problem with trace formulas is that they diverge where we need them, at the individual energy eigenvalues. What to do? Much of the quantum chaos literature responds to the challenge of wrestling the trace formulas by replacing the delta functions in the density of states (36.16) by Gaussians. But there is no need to do this - we can compute the eigenenergies without any further ado by remembering that the smart way to determine the eigenvalues of linear operators is by determining zeros of their spectral determinants.

A sensible way to compute energy levels is to construct the spectral determinant whose zeroes yield the eigenenergies, $\det(\hat{H} - E)_{sc} = 0$. A first guess might be that the spectral determinant is the Hadamard product of form

$$\det(\hat{H} - E) = \prod_n (E - E_n),$$

but this product is not well defined, since for fixed E we multiply larger and larger numbers $(E - E_n)$. This problem is dealt with by *regularization*, discussed below in appendix 39.1.2. Here we offer an impressionistic sketch of regularization.

The logarithmic derivative of $\det(\hat{H} - E)$ is the (formal) trace of the Green's function

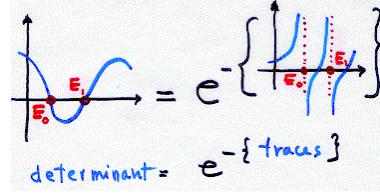
$$-\frac{d}{dE} \ln \det(\hat{H} - E) = \sum_n \frac{1}{E - E_n} = \text{Tr } G(E).$$

This quantity, not surprisingly, is divergent again. The relation, however, opens a way to derive a convergent version of $\det(\hat{H} - E)_{sc}$, by replacing the trace with the regularized trace

$$-\frac{d}{dE} \ln \det(\hat{H} - E)_{sc} = \text{Tr } G_{reg}(E).$$

The regularized trace still has $1/(E - E_n)$ poles at the semiclassical eigenenergies, poles which can be generated only if $\det(\hat{H} - E)_{sc}$ has a zero at $E = E_n$, see figure 39.3. By integrating and exponentiating we obtain

Figure 39.3: A sketch of how spectral determinants convert poles into zeros: The trace shows $1/(E - E_n)$ type singularities at the eigenenergies while the spectral determinant goes smoothly through zeroes.



$$\det(\hat{H} - E)_{sc} = \exp\left(-\int^E dE' \operatorname{Tr} G_{reg}(E')\right)$$

Now we can use (39.11) and integrate the terms coming from periodic orbits, using the relation (38.17) between the action and the period of a periodic orbit, $dS_p(E) = T_p(E)dE$, and the relation (36.19) between the density of states and the spectral staircase, $dN_{sc}(E) = d_0(E)dE$. We obtain the *semiclassical zeta function*

$$\det(\hat{H} - E)_{sc} = e^{i\pi N_{sc}(E)} \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{ir(S_p/\hbar - m_p\pi/2)}}{|\det(1 - M_p^r)|^{1/2}}\right). \quad (39.12)$$

We already know from the study of deterministic evolution operator spectra of chapter 22 that this can be evaluated by means of cycle expansions. The beauty of this formula is that everything on the right side – the cycle action S_p , the topological index m_p and monodromy matrix M_p determinant – is intrinsic, coordinate-choice independent property of the cycle p .

chapter 23

39.3 One-degree of freedom systems

It has been a long trek, a stationary phase upon stationary phase. Let us check whether the result makes sense even in the simplest case, for quantum mechanics in one spatial dimension.

In one dimension the average density of states follows from the 1-degree of freedom form of the oscillating density (39.10) and of the average density (39.8)

$$d(E) = \frac{T_p(E)}{2\pi\hbar} + \sum_r \frac{T_p(E)}{\pi\hbar} \cos(rS_p(E)/\hbar - rm_p(E)\pi/2). \quad (39.13)$$

The deterministic particle oscillates in a single potential well with period $T_p(E)$. There is no monodromy matrix to evaluate, as in one dimension there is only the parallel coordinate, and no transverse directions. The r repetition sum in (39.13) can be rewritten by using the Fourier series expansion of a delta spike train

$$\sum_{n=-\infty}^{\infty} \delta(x - n) = \sum_{k=-\infty}^{\infty} e^{i2\pi kx} = 1 + \sum_{k=1}^{\infty} 2 \cos(2\pi kx).$$

We obtain

$$d(E) = \frac{T_p(E)}{2\pi\hbar} \sum_n \delta(S_p(E)/2\pi\hbar - m_p(E)/4 - n). \quad (39.14)$$

This expression can be simplified by using the relation (38.17) between T_p and S_p , and the identity (19.7) $\delta(x - x^*) = |f'(x)|\delta(f(x))$, where x^* is the only zero of the function $f(x^*) = 0$ in the interval under consideration. We obtain

$$d(E) = \sum_n \delta(E - E_n),$$

where the energies E_n are the zeroes of the arguments of delta functions in (39.14)

$$S_p(E_n)/2\pi\hbar = n - m_p(E)/4,$$

where $m_p(E) = m_p = 2$ for smooth potential at both turning points, and $m_p(E) = m_p = 4$ for two billiard (infinite potential) walls. These are precisely the *Bohr-Sommerfeld quantized energies* E_n , defined by the condition

$$\oint p(q, E_n) dq = h \left(n - \frac{m_p}{4} \right). \quad (39.15)$$

In this way the trace formula recovers the well known 1-degree of freedom quantization rule. In one dimension, the average of states can be expressed from the quantization condition. At $E = E_n$ the exact number of states is n , while the average number of states is $n - 1/2$ since the staircase function $N(E)$ has a unit jump in this point

$$N_{sc}(E) = n - 1/2 = S_p(E)/2\pi\hbar - m_p(E)/4 - 1/2. \quad (39.16)$$

The 1-degree of freedom spectral determinant follows from (39.12) by dropping the monodromy matrix part and using (39.16)

$$\det(\hat{H} - E)_{sc} = \exp\left(-\frac{i}{2\hbar}S_p + \frac{i\pi}{2}m_p\right) \exp\left(-\sum_r \frac{1}{r} e^{\frac{i}{\hbar}rS_p - \frac{i\pi}{2}rm_p}\right). \quad (39.17)$$

Summation yields a logarithm by $\sum_r t^r/r = -\ln(1 - t)$ and we get

$$\begin{aligned} \det(\hat{H} - E)_{sc} &= e^{-\frac{i}{2\hbar}S_p + \frac{im_p}{4} + \frac{i\pi}{2}} (1 - e^{\frac{i}{\hbar}S_p - i\frac{m_p}{2}}) \\ &= 2 \sin\left(S_p(E)/\hbar - m_p(E)/4\right). \end{aligned} \quad (39.18)$$

So in one dimension, where there is only one periodic orbit for a given energy E , nothing is gained by going from the trace formula to the spectral determinant. The spectral determinant is a real function for real energies, and its zeros are again the Bohr-Sommerfeld quantized eigenenergies (39.15).

39.4 Two-degrees of freedom systems

For flows in two configuration dimensions the monodromy matrix M_p has two eigenvalues Λ_p and $1/\Lambda_p$, as explained in sect. 8.4. Isolated periodic orbits can be elliptic or hyperbolic. Here we discuss only the hyperbolic case, when the

eigenvalues are real and their absolute value is not equal to one. The determinant appearing in the trace formulas can be written in terms of the expanding eigenvalue as

$$|\det(1 - M_p^r)|^{1/2} = |\Lambda_p^r|^{1/2} \left(1 - 1/\Lambda_p^r\right),$$

and its inverse can be expanded as a geometric series

$$\frac{1}{|\det(1 - M_p^r)|^{1/2}} = \sum_{k=0}^{\infty} \frac{1}{|\Lambda_p^r|^{1/2} \Lambda_p^{kr}}.$$

With the 2-degrees of freedom expression for the average density of states (39.9) the spectral determinant becomes

$$\begin{aligned} \det(\hat{H} - E)_{sc} &= e^{\frac{i m \mathcal{A} E}{2 \hbar^2}} \exp \left(- \sum_p \sum_{r=1}^{\infty} \sum_{k=0}^{\infty} \frac{e^{i r (S_p / \hbar - m_p \pi / 2)}}{r |\Lambda_p^r|^{1/2} \Lambda_p^{kr}} \right) \\ &= e^{\frac{i m \mathcal{A} E}{2 \hbar^2}} \prod_p \prod_{k=0}^{\infty} \left(1 - \frac{e^{\frac{i}{\hbar} S_p - \frac{i \pi}{2} m_p}}{|\Lambda_p|^{1/2} \Lambda_p^k} \right). \end{aligned} \quad (39.19)$$

Résumé

Spectral determinants and dynamical zeta functions arise both in deterministic and quantum mechanics because in both the dynamical evolution can be described by the action of linear evolution operators on infinite-dimensional vector spaces. In quantum mechanics the periodic orbit theory arose from studies of semi-conductors, and the unstable periodic orbits have been measured in experiments [8] on the very paradigm of Bohr's atom, the hydrogen atom, this time in strong external fields.

In practice, most “quantum chaos” calculations take the stationary phase approximation to quantum mechanics (the Gutzwiller trace formula, possibly improved by including tunneling periodic trajectories, diffraction corrections, etc.) as the point of departure. Once the stationary phase approximation is made, what follows is *deterministic* in the sense that all quantities used in periodic orbit calculations - actions, stabilities, geometrical phases - are deterministic quantities. The problem is then to understand and control the convergence of deterministic periodic orbit formulas.

While various periodic orbit formulas are formally equivalent, practice shows that some are preferable to others. Three classes of periodic orbit formulas are frequently used:

Trace formulas. The trace of the semiclassical Green's function

$$\text{Tr } G_{sc}(E) = \int dq G_{sc}(q, q, E)$$

is given by a sum over the periodic orbits (39.11). While easiest to derive, in calculations the trace formulas are inconvenient for anything other than the leading eigenvalue estimates, as they tend to be divergent in the region of physical interest. In deterministic dynamics trace formulas hide under a variety of appellations such as the f - α or multifractal formalism; in quantum mechanics they are known as the Gutzwiller trace formulas.

Zeros of Ruelle or dynamical zeta functions

$$1/\zeta(s) = \prod_p (1 - t_p), \quad t_p = \frac{1}{|\Lambda_p|^{1/2}} e^{\frac{i}{\hbar} S_p - i\pi m_p/2}$$

yield, in combination with cycle expansions, the semiclassical estimates of *quantum* resonances. For hyperbolic systems the dynamical zeta functions have good convergence and are a useful tool for determination of deterministic and quantum mechanical averages.

Spectral determinants, Selberg-type zeta functions, Fredholm determinants, functional determinants are the natural objects for spectral calculations, with convergence better than for dynamical zeta functions, but with less transparent cycle expansions. The 2-degrees of freedom semiclassical spectral determinant (39.19)

$$\det(\hat{H} - E)_{sc} = e^{i\pi N_{sc}(E)} \prod_p \prod_{k=0}^{\infty} \left(1 - \frac{e^{iS_p/\hbar - i\pi m_p/2}}{|\Lambda_p|^{1/2} \Lambda_p^k} \right)$$

is a typical example. Most periodic orbit calculations are based on cycle expansions of such determinants.

As we have assumed repeatedly during the derivation of the trace formula that the periodic orbits are isolated, and do not form families (as is the case for integrable systems or in KAM tori of systems with mixed phase space), the formulas discussed so far are valid only for hyperbolic and elliptic periodic orbits.

For deterministic dynamical flows and number theory, spectral determinants and zeta functions are exact. The quantum-mechanical ones, derived by the Gutzwiller approach, are at best only the stationary phase approximations to the exact quantum spectral determinants, and for quantum mechanics an important conceptual problem arises already at the level of derivation of the semiclassical formulas; how accurate are they, and can the periodic orbit theory be systematically improved?

Commentary

Remark 39.1. Gutzwiller quantization of deterministically chaotic systems. The derivation given here and in sects. 38.3 and 39.1 follows closely the excellent exposition [6] by Martin Gutzwiller, the inventor of the trace formula. The derivation presented here is self contained, but refs. [11, 12] might also be of help to the student.

Remark 39.2. Cat map quantization. (Continued from remark 8.3) Hannay and Berry [7] *Quantization of linear maps on a torus – Fresnel diffraction by a periodic grating* might be the first to use Arnol’d cat map [1] in the quantization context. They exhibit various periodic orbits (cycles) and rules for constructing them. An extensive analysis of the use of the periodic orbits of the cat map in quantization has been performed by Keating [10]. Jin and Zworski [9] write down the deterministic trace formula for Anosov flows. Creagh [3] studies semiclassical approximations to the spectra of smoothly perturbed quantum cat maps. Boasman and Keating [2] study semiclassical asymptotics of perturbed cat maps. García-Mata and Saraceno [5] study the Arnol’d cat map with a small sinusoidal perturbation and note that its Lyapunov exponent $\lambda = \ln[(3 + \sqrt{5})/2]$ is almost independent of the weak nonlinearity, while the Ruelle-Pollicot resonances are very sensitive to it. Rivas [13] derives an exact expression for a (linear) cat map. Faure [4] investigates resonances of a “prequantum” cat map.

Remark 39.3. Zeta functions. For “zeta function” nomenclature, see remark 22.4 on page 441.

References

- [1] V. I. Arnol’d and A. Avez, *Ergodic Problems of Classical Mechanics* (Addison-Wesley, Redwood City, 1989).
- [2] P. A. Boasman and J. P. Keating, “Semiclassical asymptotics of perturbed cat maps”, *Proc. Roy. Soc. Ser A* **449**, 629–653 (1995).
- [3] S. C. Creagh, “Quantum zeta function for perturbed cat maps”, *Chaos* **5**, 477–493 (1995).
- [4] F. Faure, “Prequantum chaos: Resonances of the prequantum cat map”, *J. Modern Dynamics* **1**, 255–285 (2007).
- [5] I. García-Mata and M. Saraceno, “Spectral properties and classical decays in quantum open systems”, *Phys. Rev. E* **69**, 056211 (2004).
- [6] M. C. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer, New York, 1990).
- [7] J. H. Hannay and M. V. Berry, “Quantization of linear maps on a torus – Fresnel diffraction by a periodic grating”, *Physica D* **1**, 267–290 (1980).
- [8] A. Holle, J. Main, G. Wiebusch, H. Rottke, and K. H. Welge, “Quasi-Landau spectrum of the chaotic diamagnetic hydrogen atom”, *Phys. Rev. Lett.* **61**, 161–164 (1988).
- [9] L. Jin and M. Zworski, “A local trace formula for Anosov flows”, *Ann. Inst. H. Poincaré* **18**, 1–35 (2016).
- [10] J. P. Keating, “The cat maps: quantum mechanics and classical motion”, *Nonlinearity* **4**, 309–341 (1991).
- [11] R. G. Littlejohn, “The Van Vleck formula, Maslov theory, and phase space geometry”, *J. Stat. Phys.* **68**, 7–50 (1992).
- [12] L. Reichl, *The Transition to Chaos: Conservative Classical Systems and Quantum Manifestations*, 2nd ed. (Springer, New York, 2004).

- [13] A. M. F. Rivas, “Semiclassical matrix elements for a chaotic propagator in the scar function basis”, J. Phys. A **46**, 145101 (2013).

39.5 Exercises

39.1. **Monodromy matrix from second variations of the action.** Show that

$$D_{\perp j}/D'_{\perp j} = (1 - M) \quad (39.20)$$

39.2. **Volume of d -dimensional sphere.** Show that the volume of a d -dimensional sphere of radius r equals $\pi^{d/2} r^d / \Gamma(1 + d/2)$. Show that $\Gamma(1 + d/2) = \Gamma(d/2)d/2$.

39.3. **Average density of states in 1 dimension.** Show that in one dimension the average density of states is given by (39.8)

$$\bar{d}(E) = \frac{T(E)}{2\pi\hbar},$$

where $T(E)$ is the time period of the 1-dimensional motion and show that

$$\bar{N}(E) = \frac{S(E)}{2\pi\hbar}, \quad (39.21)$$

where $S(E) = \oint p(q, E) dq$ is the action of the orbit.

39.4. **Average density of states in 2 dimensions.** Show that in 2 dimensions the average density of states is given by (39.9)

$$\bar{d}(E) = \frac{m\mathcal{A}(E)}{2\pi\hbar^2},$$

where $\mathcal{A}(E)$ is the deterministically allowed area of configuration space for which $U(q) < E$.

Chapter 40

Quantum scattering

Scattering is easier than gathering.

—Irish proverb

(A. Wirzba, P. Cvitanović and N. Whelan)

SO FAR the trace formulas have been derived assuming that the system under consideration is bound. As we shall now see, we are in luck - the semiclassics of bound systems is all we need to understand the semiclassics for open, scattering systems as well. We start by a brief review of the quantum theory of elastic scattering of a point particle from a (repulsive) potential, and then develop the connection to the standard Gutzwiller theory for bound systems. We do this in two steps - first, a heuristic derivation which helps us understand in what sense density of states is “density,” and then we sketch a general derivation of the central result of the spectral theory of quantum scattering, the Krein-Friedel-Lloyd formula. The end result is that we establish a connection between the scattering resonances (both positions and widths) of an open quantum system and the poles of the trace of the Green’s function, which we learned to analyze in earlier chapters.

40.1 Density of states

For a scattering problem the density of states (36.16) appear ill defined since formulas such as (39.6) involve integration over infinite spatial extent. What we will now show is that a quantity that makes sense physically is the difference of two densities - the first with the scatterer present and the second with the scatterer absent.

In non-relativistic dynamics the relative motion can be separated from the center-of-mass motion. Therefore the elastic scattering of two particles can be treated as the scattering of one particle from a static potential $V(q)$. We will study

the scattering of a point-particle of (reduced) mass m by a short-range potential $V(q)$, excluding *inter alia* the Coulomb potential. (The Coulomb potential decays slowly as a function of q so that various asymptotic approximations which apply to general potentials fail for it.) Although we can choose the spatial coordinate frame freely, it is advisable to place its origin somewhere near the geometrical center of the potential. The scattering problem is solved, if a scattering solution to the time-independent Schrödinger equation (37.2)

$$\left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial q^2} + V(q) \right) \phi_{\vec{k}}(q) = E \phi_{\vec{k}}(q) \quad (40.1)$$

can be constructed. Here E is the energy, $\vec{p} = \hbar \vec{k}$ the initial momentum of the particle, and $\vec{k}$ the corresponding wave vector.

When the argument $r = |q|$ of the wave function is large compared to the typical size a of the scattering region, the Schrödinger equation effectively becomes a free particle equation because of the short-range nature of the potential. In the asymptotic domain $r \gg a$, the solution $\phi_{\vec{k}}(q)$ of (40.1) can be written as superposition of ingoing and outgoing solutions of the free particle Schrödinger equation for fixed angular momentum:

$$\phi(q) = A\phi^{(-)}(q) + B\phi^{(+)}(q), \quad (+ \text{ boundary conditions}),$$

where in 1-dimensional problems $\phi^{(-)}(q)$, $\phi^{(+)}(q)$ are the “left,” “right” moving plane waves, and in higher-dimensional scattering problems the “incoming,” “outgoing” radial waves, with the constant matrices A , B fixed by the boundary conditions. What are the boundary conditions? The scatterer can modify only the outgoing waves (see figure 40.1), since the incoming ones, by definition, have yet to encounter the scattering region. This defines the quantum mechanical scattering matrix, or the *S matrix*

$$\phi_m(r) = \phi_m^{(-)}(r) + S_{mm'} \phi_{m'}^{(+)}(r). \quad (40.2)$$

All scattering effects are incorporated in the deviation of $\mathbf{S}$ from the unit matrix, the transition matrix $\mathbf{T}$

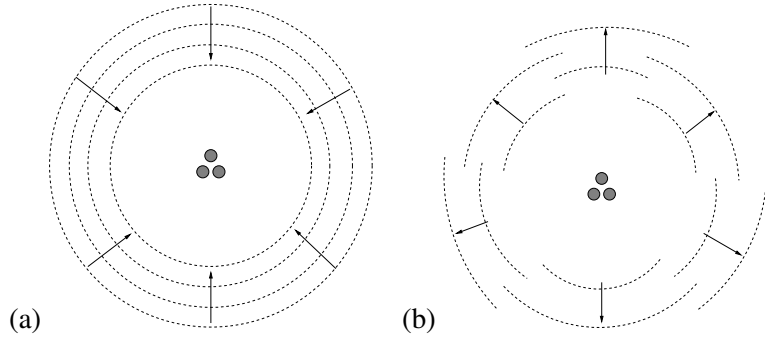
$$\mathbf{S} = \mathbf{1} - i\mathbf{T}. \quad (40.3)$$

For concreteness, we have specialized to two dimensions, although the final formula is true for arbitrary dimensions. The indices m and m' are the angular momenta quantum numbers for the incoming and outgoing state of the scattering wave function, labeling the S -matrix elements $S_{mm'}$. More generally, given a set of quantum numbers β, γ , the S matrix is a collection $S_{\beta\gamma}$ of transition amplitudes $\beta \rightarrow \gamma$ normalized such that $|S_{\beta\gamma}|^2$ is the probability of the $\beta \rightarrow \gamma$ transition. The total probability that the ingoing state β ends up in some outgoing state must add up to unity

$$\sum_{\gamma} |S_{\beta\gamma}|^2 = 1, \quad (40.4)$$

so the S matrix is unitary: $\mathbf{S}^\dagger \mathbf{S} = \mathbf{S} \mathbf{S}^\dagger = \mathbf{1}$.

Figure 40.1: (a) Incoming spherical waves running into an obstacle. (b) Superposition of outgoing spherical waves scattered from an obstacle.



We have already encountered a solution to the 2-dimensional problem; free particle propagation Green's function (38.48) is a radial solution, given in terms of the Hankel function

$$G_0(r, 0, E) = -\frac{im}{2\hbar^2} H_0^{(+)}(kr),$$

where we have used $S_0(r, 0, E)/\hbar = kr$ for the action. The m th angular momentum eigenfunction is proportional to $\phi_m^{(\pm)}(q) \propto H_m^{(\pm)}(kr)$, and given a potential $V(q)$ we can in principle compute the infinity of matrix elements $S_{mm'}$. We will not need much information about $H_m^{(t)}(kr)$, other than that for large r its asymptotic form is

$$H^\pm \propto e^{\pm ikr}$$

In general, the potential $V(q)$ is not radially symmetric and (40.1) has to be solved numerically, by explicit integration, or by diagonalizing a large matrix in a specific basis. To simplify things a bit, we assume for the time being that a radially symmetric scatterer is centered at the origin; the final formula will be true for arbitrary asymmetric potentials. Then the solutions of the Schrödinger equation (37.2) are separable, $\phi_m(q) = \phi(r)e^{im\theta}$, $r = |q|$, the scattering matrix cannot mix different angular momentum eigenstates, and S is diagonal in the radial basis (40.2) with matrix elements given by

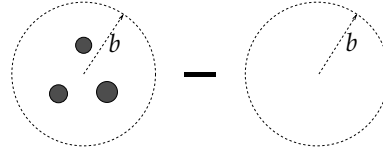
$$S_m(k) = e^{2i\delta_m(k)}. \quad (40.5)$$

The matrix is unitary so in a diagonal basis all entries are pure phases. This means that an incoming state of the form $H_m^{(-)}(kr)e^{im\theta}$ gets scattered into an outgoing state of the form $S_m(k)H_m^{(+)}(kr)e^{im\theta}$, where $H_m^{(\mp)}(z)$ are incoming and outgoing Hankel functions respectively. We now embed the scatterer in a infinite cylindrical well of radius R , and will later take $R \rightarrow \infty$. Angular momentum is still conserved so that each eigenstate of this (now bound) problem corresponds to some value of m . For large $r \gg a$ each eigenstate is of the asymptotically free form

$$\begin{aligned} \phi_m(r) &\approx e^{im\theta} (S_m(k)H_m^{(+)}(kr) + H_m^{(-)}(kr)) \\ &\approx \cdots \cos(kr + \delta_m(k) - \chi_m), \end{aligned} \quad (40.6)$$

where $\cdots$ is a common prefactor, and $\chi_m = m\pi/2 + \pi/4$ is an annoying phase factor from the asymptotic expansion of the Hankel functions that will play no role in what follows.

Figure 40.2: The “difference” of two bounded reference systems, one with and one without the scattering system.



The state (40.6) must satisfy the external boundary condition that it vanish at $r = R$. This implies the quantization condition

$$k_n R + \delta_m(k_n) - \chi_m = \pi(n + 1/2).$$

We now ask for the difference in the eigenvalues of two consecutive states of fixed m . Since R is large, the density of states is high, and the phase $\delta_m(k)$ does not change much over such a small interval. Therefore, to leading order we can include the effect of the change of the phase on state $n + 1$ by Taylor expanding,

$$k_{n+1} R + \delta_m(k_{n+1}) - \chi_m \approx \pi(n + 1/2) + (k_{n+1} - k_n) \delta'_m(k_n) R.$$

Taking the difference of the two equations we obtain $\Delta k \approx \pi(R + \delta'_m(k))^{-1}$. This is the eigenvalue spacing which we now interpret as the inverse of the density of states within m angular momentum subspace

$$d_m(k) \approx \frac{1}{\pi} (R + \delta'_m(k)).$$

The R term is essentially the $1 - d$ Weyl term (39.8), appropriate to $1 - d$ radial quantization. For large R , the dominant behavior is given by the size of the circular enclosure with a correction in terms of the derivative of the scattering phase shift, approximation accurate to order $1/R$. However, not all is well; the area under consideration tends to infinity. We regularize this by subtracting from the result from the free particle density of states $d_0(k)$, for the same size container, but this time without any scatterer, figure 40.2. We also sum over all m values so that

$$\begin{aligned} d(k) - d_0(k) &= \frac{1}{\pi} \sum_m \delta'_m(k) = \frac{1}{2\pi i} \sum_m \frac{d}{dk} \log S_m \\ &= \frac{1}{2\pi i} \text{Tr} \left(S^\dagger \frac{dS}{dk} \right). \end{aligned} \quad (40.7)$$

The first line follows from the definition of the phase shifts (40.5) while the second line follows from the unitarity of S so that $S^{-1} = S^\dagger$. We can now take the limit $R \rightarrow \infty$ since the R dependence has been cancelled away.

This is essentially what we want to prove since for the left hand side we already have the semiclassical theory for the trace of the difference of Green's functions,

$$d(k) - d_0(k) = -\frac{1}{2\pi k} \text{Im}(\text{tr}(G(k) - G_0(k))). \quad (40.8)$$

There are a number of generalizations. This can be done in any number of dimensions. It is also more common to do this as a function of energy and not

wave number k . However, as the asymptotic dynamics is free wave dynamics labeled by the wavenumber k , we have adapted k as the natural variable in the above discussion.

Finally, we state without proof that the relation (40.7) applies even when there is no circular symmetry. The proof is more difficult since one cannot appeal to the phase shifts δ_m but must work directly with a non-diagonal S matrix.

40.2 Quantum mechanical scattering matrix

The results of the previous section indicate that there is a connection between the scattering matrix and the trace of the quantum Green's function (more formally between the difference of the Green's function with and without the scattering center.) We now show how this connection can be derived in a more rigorous manner. We will also work in terms of the energy E rather than the wavenumber k , since this is the more usual exposition. Suppose particles interact via forces of sufficiently short range, so that in the remote past they were in a free particle state labeled β , and in the distant future they will likewise be free, in a state labeled γ . In the Heisenberg picture the S -matrix is defined as $\mathbf{S} = \Omega_- \Omega_+^\dagger$ in terms of the Møller operators

$$\Omega_\pm = \lim_{t \rightarrow \pm\infty} e^{iHt/\hbar} e^{-iH_0 t/\hbar}, \quad (40.9)$$

where H is the full Hamiltonian, whereas H_0 is the free Hamiltonian. In the interaction picture the S -matrix is given by

$$\begin{aligned} \mathbf{S} &= \Omega_+^\dagger \Omega_- = \lim_{t \rightarrow \infty} e^{iH_0 t/\hbar} e^{-2iHt/\hbar} e^{iH_0 t/\hbar} \\ &= T \exp \left(-i \int_{-\infty}^{+\infty} dt H'(t) \right), \end{aligned} \quad (40.10)$$

where $H' = V = H - H_0$ is the interaction Hamiltonian and T is the time-ordering operator. In stationary scattering theory the S matrix has the following spectral representation

$$\begin{aligned} S &= \int_0^\infty dE S(E) \delta(H_0 - E) \\ S(E) &= Q_+(E) Q_-^{-1}(E), \quad Q_\pm(E) = \mathbf{1} + (H_0 - E \pm i\epsilon)^{-1} V, \end{aligned} \quad (40.11)$$

such that

$$\text{Tr} \left[S^\dagger(E) \frac{d}{dE} S(E) \right] = \text{Tr} \left[\frac{1}{H_0 - E - i\epsilon} - \frac{1}{H - E - i\epsilon} - (\epsilon \leftrightarrow -\epsilon) \right]. \quad (40.12)$$

The manipulations leading to (40.12) are justified if the operators $Q_\pm(E)$ can be linked to trace-class operators.

appendix A40

We can now use this result to derive the Krein-Lloyd formula which is the central result of this chapter. The Krein-Lloyd formula provides the connection between the trace of the Green's function and the poles of the scattering matrix, implicit in all of the trace formulas for open quantum systems which will be presented in the subsequent chapters.

40.3 Krein-Friedel-Lloyd formula

The link between quantum mechanics and semiclassics for scattering problems is provided by the semiclassical limit of the Krein-Friedel-Lloyd sum for the spectral density which we now derive. This derivation builds on the results of the last section and extends the discussion of the opening section.

In chapter 38 we linked the spectral density (see (36.16)) of a bounded system

$$d(E) \equiv \sum_n \delta(E_n - E) \quad (40.13)$$

via the identity

$$\begin{aligned} \delta(E_n - E) &= -\lim_{\epsilon \rightarrow 0} \frac{1}{\pi} \operatorname{Im} \frac{1}{E - E_n + i\epsilon} \\ &= -\lim_{\epsilon \rightarrow 0} \frac{1}{\pi} \operatorname{Im} \langle E_n | \frac{1}{E - H + i\epsilon} | E_n \rangle \\ &= \frac{1}{2\pi i} \lim_{\epsilon \rightarrow 0} \left\langle E_n \left| \frac{1}{E - H - i\epsilon} - \frac{1}{E - H + i\epsilon} \right| E_n \right\rangle \end{aligned} \quad (40.14)$$

to the trace of the Green's function (39.1.1). Furthermore, in the semiclassical approximation, the trace of the Green's function is given by the Gutzwiller trace formula (39.11) in terms of a smooth Weyl term and an oscillating contribution of periodic orbits.

Therefore, the task of constructing the semiclassics of a scattering system is completed, if we can find a connection between the spectral density $d(E)$ and the scattering matrix S . We will see that (40.12) provides the clue. Note that the right hand side of (40.12) has nearly the structure of (40.14) when the latter is inserted into (40.13). The principal difference between these two types of equations is that the S matrix refers to *outgoing* scattering wave functions which are not normalizable and which have a *continuous* spectrum, whereas the spectral density $d(E)$ refers to a bound system with normalizable wave functions with a discrete spectrum. Furthermore, the bound system is characterized by a *hermitian* operator, the Hamiltonian H , whereas the scattering system is characterized by a *unitary* operator, the S -matrix. How can we reconcile these completely different classes of wave functions, operators and spectra? The trick is to put our scattering system into a finite box as in the opening section. We choose a spherical container with

radius R and with its center at the center of our finite scattering system. Our scattering potential $V(\vec{r})$ will be unaltered within the box, whereas at the box walls we will choose an infinitely high potential, with the Dirichlet boundary conditions at the outside of the box:

$$\phi(\vec{r})|_{r=R} = 0. \quad (40.15)$$

In this way, for any finite value of the radius R of the box, we have mapped our scattering system into a bound system with a spectral density $d(E; R)$ over discrete eigenenergies $E_n(R)$. It is therefore important that our scattering potential was chosen to be short-ranged to start with. (Which explains why the Coulomb potential requires special care.) The hope is that in the limit $R \rightarrow \infty$ we will recover the scattering system. But some care is required in implementing this. The smooth Weyl term $\bar{d}(E; R)$ belonging to our box with the enclosed potential V diverges for a spherical 2-dimensional box of radius R quadratically, as $\pi R^2/(4\pi)$ or as R^3 in the 3-dimensional case. This problem can easily be cured if the spectral density of an empty reference box of the *same* size (radius R) is subtracted (see figure 40.2). Then all the divergences linked to the increasing radius R in the limit $R \rightarrow \infty$ drop out of the difference. Furthermore, in the limit $R \rightarrow \infty$ the energy-eigenfunctions of the box are only normalizable as a delta distribution, similarly to a plane wave. So we seem to recover a continuous spectrum. Still the problem remains that the wave functions do not discriminate between incoming and outgoing waves, whereas this symmetry, namely the hermiticity, is broken in the scattering problem. The last problem can be tackled if we replace the spectral density over discrete delta distributions by a smoothed spectral density with a small finite imaginary part η in the energy E :

$$d(E + i\eta; R) \equiv \frac{1}{i2\pi} \sum_n \left\{ \frac{1}{E - E_n(R) - i\eta} - \frac{1}{E - E_n(R) + i\eta} \right\}. \quad (40.16)$$

Note that $d(E + i\eta; R) \neq d(E - i\eta; R) = -d(E + i\eta; R)$. By the introduction of the positive *finite* imaginary part η the time-dependent behavior of the wave function has effectively been altered from an oscillating one to a decaying one and the hermiticity of the Hamiltonian is removed. Finally the limit $\eta \rightarrow 0$ can be carried out, respecting the order of the limiting procedures. First, the limit $R \rightarrow \infty$ has to be performed for a *finite* value of η , only then the limit $\eta \rightarrow 0$ is allowed. In practice, one can try to work with a finite value of R , but then it will turn out (see below) that the scattering system is only recovered if $R\sqrt{\eta} \gg 1$.

Let us summarize the relation between the smoothed spectral densities $d(E + i\eta; R)$ of the boxed potential and $d^{(0)}(E + i\eta; R)$ of the empty reference system and the S matrix of the corresponding scattering system:

$$\begin{aligned} \lim_{\eta \rightarrow +0} \lim_{R \rightarrow \infty} (d(E + i\eta; R) - d^{(0)}(E + i\eta; R)) &= \frac{1}{2\pi i} \text{Tr} \left[S^\dagger(E) \frac{d}{dE} S(E) \right] \\ &= \frac{1}{2\pi i} \text{Tr} \frac{d}{dE} \ln S(E) = \frac{1}{2\pi i} \frac{d}{dE} \ln \det S(E). \end{aligned} \quad (40.17)$$

This is the *Krein-Friedel-Lloyd formula*. It replaces the scattering problem by the difference of two bounded reference billiards of the same radius R which finally will be taken to infinity. The first billiard contains the scattering region or

potentials, whereas the other does not (see figure 40.2). Here $d(E + i\eta; R)$ and $d^{(0)}(E + i\eta; R)$ are the *smoothed* spectral densities in the presence or in the absence of the scatterers, respectively. In the semiclassical approximation, they are replaced by a Weyl term (39.10) and an oscillating sum over periodic orbits. As in (39.2), the trace formula (40.17) can be integrated to give a relation between the smoothed staircase functions and the determinant of the S -matrix:

$$\lim_{\eta \rightarrow +0} \lim_{R \rightarrow \infty} (N(E+i\eta; R) - N^{(0)}(E+i\eta; R)) = \frac{1}{2\pi i} \ln \det S(E). \quad (40.18)$$

Furthermore, in both versions of the Krein-Friedel-Lloyd formulas the energy argument $E + i\eta$ can be replaced by the wavenumber argument $k + i\eta'$. These expressions only make sense for wavenumbers on or above the real k -axis. In particular, if k is chosen to be real, η' must be greater than zero. Otherwise, the exact left hand sides (40.18) and (40.17) would give discontinuous staircase or even delta function sums, respectively, whereas the right hand sides are continuous to start with, since they can be expressed by continuous phase shifts. Thus the order of the two limits in (40.18) and (40.17) is essential.

The necessity of the $+i\eta$ prescription can also be understood by purely phenomenological considerations in the semiclassical approximation: Without the $i\eta$ term there is no reason why one should be able to neglect spurious periodic orbits which are there solely because of the introduction of the confining boundary. The subtraction of the second (empty) reference system removes those spurious periodic orbits which never encounter the scattering region – in addition to the removal of the divergent Weyl term contributions in the limit $R \rightarrow \infty$. The periodic orbits that encounter both the scattering region and the external wall would still survive the first limit $R \rightarrow \infty$, if they were not exponentially suppressed by the $+i\eta$ term because of their

$$e^{iL(R)\sqrt{2m(E+i\eta)}} = e^{iL(R)k} e^{-L(R)\eta'}$$

behavior. As the length $L(R)$ of a spurious periodic orbit grows linearly with the radius R . The bound $R\eta' \gg 1$ is an essential precondition on the suppression of the unwanted spurious contributions of the container if the Krein-Friedel-Lloyd formulas (40.17) and (40.18) are evaluated at a finite value of R .

exercise 40.1

Finally, the semiclassical approximation can also help us in the interpretation of the Weyl term contributions for scattering problems. In scattering problems the Weyl term appears with a negative sign. The reason is the subtraction of the empty container from the container with the potential. If the potential is a dispersing billiard system (or a finite collection of dispersing billiards), we expect an excluded volume (or the sum of excluded volumes) relative to the empty container. In other words, the Weyl term contribution of the empty container is larger than of the filled one and therefore a negative net contribution is left over. Second, if the scattering potential is a collection of a finite number of non-overlapping scattering regions, the Krein-Friedel-Lloyd formulas show that the corresponding Weyl contributions are completely independent of the position of the single scatterers, as long as these do not overlap.

40.4 Wigner time delay

The term $\frac{d}{dE} \ln \det S$ in the density formula (40.17) is dimensionally time. This suggests another, physically important interpretation of such formulas for scattering systems, the Wigner delay, defined as

$$\begin{aligned} d(k) &= \frac{d}{dk} \text{Argdet}(\mathbf{S}(k)) \\ &= -i \frac{d}{dk} \log \det(\mathbf{S}(k)) \\ &= -i \text{Tr} \left(\mathbf{S}^\dagger(k) \frac{d\mathbf{S}}{dk}(k) \right) \end{aligned} \quad (40.19)$$

and can be shown to equal the total delay of a wave packet in a scattering system. We now review this fact.

A related quantity is the total scattering *phase shift* $\Theta(k)$ defined as

$$\det \mathbf{S}(k) = e^{+i\Theta(k)},$$

so that $d(k) = \frac{d}{dk} \Theta(k)$.

The time delay may be both positive and negative, reflecting attractive respectively repulsive features of the scattering system. To elucidate the connection between the scattering determinant and the time delay we study a plane wave:

The phase of a wave packet will have the form:

$$\phi = \vec{k} \cdot \vec{x} - \omega t + \Theta.$$

The center of the wave packet will be determined by the principle of stationary phase:

$$0 = d\phi = d\vec{k} \cdot \vec{x} - d\omega t + d\Theta.$$

Hence the packet is located at

$$\vec{x} = \frac{\partial \omega}{\partial \vec{k}} t - \frac{\partial \Theta}{\partial \vec{k}}.$$

The first term is just the group velocity times the given time t . Thus the packet is retarded by a length given by the derivative of the phase shift with respect to the wave vector $\vec{k}$. The arrival of the wave packet at the position $\vec{x}$ will therefore be delayed. This *time delay* can similarly be found as

$$\tau(\omega) = \frac{\partial \Theta(\omega)}{\partial \omega}.$$

To show this we introduce the *slowness* of the phase $\vec{s} = \vec{k}/\omega$ for which $\vec{s} \cdot \vec{v}_g = 1$, where $\vec{v}_g$ is the group velocity to get

$$d\vec{k} \cdot \vec{x} = \vec{s} \cdot \vec{x} d\omega = \frac{x}{v_g} d\omega,$$

since we may assume $\vec{x}$ is parallel to the group velocity (consistent with the above). Hence the arrival time becomes

$$t = \frac{x}{v_g} + \frac{\partial \Theta(\omega)}{\partial \omega}.$$

If the scattering matrix is not diagonal, one interprets

$$\Delta t_{ij} = \text{Re} \left(-i S_{ij}^{-1} \frac{\partial S_{ij}}{\partial \omega} \right) = \text{Re} \left(\frac{\partial \Theta_{ij}}{\partial \omega} \right)$$

as the delay in the j th scattering channel after an injection in the i th. The probability for appearing in channel j goes as $|S_{ij}|^2$ and therefore the average delay for the incoming states in channel i is

$$\begin{aligned} \langle \Delta t_i \rangle &= \sum_j |S_{ij}|^2 \Delta t_{ij} = \text{Re} \left(-i \sum_j S_{ij}^* \frac{\partial S_{ij}}{\partial \omega} \right) = \text{Re} \left(-i \mathbf{S}^\dagger \cdot \frac{\partial \mathbf{S}}{\partial \omega} \right)_{ii} \\ &= -i \left(\mathbf{S}^\dagger \cdot \frac{\partial \mathbf{S}}{\partial \omega} \right)_{ii}, \end{aligned}$$

where we have used the derivative, $\partial/\partial\omega$, of the unitarity relation $\mathbf{S} \cdot \mathbf{S}^\dagger = \mathbf{1}$ valid for real frequencies. This discussion can in particular be made for wave packets related to partial waves and superpositions of these like an incoming plane wave corresponding to free motion. The total Wigner delay therefore corresponds to the sum over all channel delays (40.19).

Commentary

Remark 40.1. Krein-Friedel-Lloyd formula. The third volume of Thirring [18], sections 3.6.14 (Levison Theorem) and 3.6.15 (the proof), or P. Scherer's thesis [15] (appendix) discusses the Levison Theorem. It helps to start with a toy example or simplified example instead of the general theorem, namely for the radially symmetric potential in a symmetric cavity. Have a look at Huang [8] chapter 10 (on the "second virial coefficient"), or Beth and Uhlenbeck [2], or Friedel [4, 5]. These results for the correction to the density of states are particular cases of the Krein formula [9, 10]. The Krein-Friedel-Lloyd formula (40.17) was derived in refs. [4, 5, 9, 10, 12, 13], see also refs. [1, 6, 7, 14–16]. The original papers are by Krein and Birman [3, 9–11] but beware, they are mathematicians. Also, have a look at pages 15–18 of Wirzba's talk on the Casimir effect [19]. Page 16 discusses the Beth-Uhlenbeck formula [2], the predecessor of the more general Krein formula for spherical cases.

Remark 40.2. Weyl term for empty container. For a discussion of why the Weyl term contribution of the empty container is larger than of the filled one and therefore a negative net contribution is left over, see ref. [15].

Remark 40.3. Wigner time delay. Wigner time delay and the Wigner-Smith time delay matrix, are powerful concepts for a statistical description of scattering. The diagonal

elements Q_{aa} of the lifetime matrix $\mathbf{Q} = -i\mathbf{S}^{-1}\partial\mathbf{S}/\partial\omega$, where $\mathbf{S}$ is the $[2N \times 2N]$ scattering matrix, are interpreted in terms of the time spent in the scattering region by a wave packet incident in one channel. As shown by Smith [17], they are the sum over all output channels (both in reflection and transmission) of $\Delta t_{ab} = \text{Re} [(-i/S_{ab})(\partial S_{ab}/\partial\omega)]$ weighted by the probability of emerging from that channel. The sum of the Q_{aa} over all $2N$ channels is the Wigner time delay $\tau_W = \sum_a Q_{aa}$, which is the trace of the lifetime matrix and is proportional to the density of states.

References

- [1] R. Balian and C. Bloch, “Solution of the Schrödinger equation in terms of classical paths”, *Ann. Phys.* **85**, 514–545 (1974).
- [2] E. Beth and G. E. Uhlenbeck, “The quantum theory of the non-ideal gas. II. Behaviour at low temperatures”, *Physica* **4**, 915–924 (1937).
- [3] M. S. Birman and D. R. Yafaev, “The spectral shift function. The papers of M. G. Krein and their further development”, *Algebra i Analiz* **4**, 1–44 (1992).
- [4] J. Friedel, “XIV. The distribution of electrons round impurities in monovalent metals”, *Phil. Mag.* **43**, 153–189 (1952).
- [5] J. Friedel, “Metallic alloys”, *Nuovo Cim.* **7**, 287–311 (1958).
- [6] P. Gaspard, Scattering resonances: classical and quantum dynamics, in *Quantum Chaos*, Vol. 19, edited by G. Casati, I. Guarneri, and U. Smilansky, Proc. Int. School of Physics Enrico Fermi (1993).
- [7] P. Gaspard and S. A. Rice, “Semiclassical quantization of the scattering from a classically chaotic repeller”, *J. Chem. Phys.* **90**, 2242–2254 (1989).
- [8] K. Huang, *Statistical Mechanics*, 2nd ed. (Wiley, New York, 1987).
- [9] M. G. Krein, “On the trace formula in perturbation theory”, *Mat. Sb. (N.S.)* **33**, 597–626 (1953).
- [10] M. G. Krein, “Perturbation determinants and formula for traces of unitary and self-adjoint operators”, *Sov. Math. Dokl.* **3**, 707–710 (1962).
- [11] M. G. Krein and M. S. Birman, “On the theory of wave operators and scattering operators”, *Dokl. Akad. Nauk SSSR* **144**, 740–744 (1962).
- [12] P. Lloyd, “Wave propagation through an assembly of spheres: II. The density of single-particle eigenstates”, *Proc. Phys. Soc.* **90**, 207–216 (1967).
- [13] P. Lloyd and P. V. Smith, “Multiple scattering theory in condensed materials”, *Adv. Phys.* **21**, 69–142 (1972).
- [14] A. Norcliffe and I. C. Percival, “Correspondence identities: I”, *J. Phys. B* **1**, 774 (1968).
- [15] P. Scherer, Quantenzustände eines Klassisch Chaotischen Billards, PhD thesis (Univ. Köln, 1991).
- [16] L. Schulman, “A path integral for spin”, *Phys. Rev.* **176**, 1558–1569 (1968).

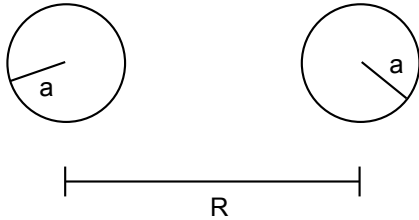
- [17] F. T. Smith, “Lifetime matrix in collision theory”, *Phys. Rev.* **118**, 349–356 (1960).
- [18] W. Thirring, *Quantum Mechanics of Atoms and Molecules*, Vol. 3, Page 48, Sects. 2.3.21 and 2.3.22 . (Springer, New York, 1979).
- [19] A. Wirzba, *A force from nothing into nothing: Casimir interactions*, Overheads, 2003.

Exercises

- 40.1. **Spurious orbits under the Krein-Friedel-Lloyd construction.** Draw examples for the three types of period orbits under the Krein-Friedel-Lloyd construction: (a) the genuine periodic orbits of the scattering region, (b) spurious periodic orbits which can be removed by the subtraction of the reference system, (c) spurious periodic orbits which cannot be removed by this subtraction. What is the role of the double limit $\eta \rightarrow 0$, container size $b \rightarrow \infty$?
- 40.2. **The one-disk scattering wave function.** Derive the one-disk scattering wave function.
(Andreas Wirzba)
- 40.3. **Quantum two-disk scattering.** Compute the quasiclassical spectral determinant

$$Z(\varepsilon) = \prod_{p,j,l} \left(1 - \frac{t_p}{\Lambda_p^{j+2l}} \right)^{j+1}$$

for the two disk problem. Use the geometry



The full quantum mechanical version of this problem can be solved by finding the zeros in k for the determinant of the matrix

$$M_{m,n} = \delta_{m,n} + \frac{(-1)^n}{2} \frac{J_m(ka)}{H_n^{(1)}(ka)} \left(H_{m-n}^{(1)}(kR) + (-1)^n H_{m+n}^{(1)}(kR) \right),$$

where J_n is the n th Bessel function and $H_n^{(1)}$ is the Hankel function of the first kind. Find the zeros of the determinant closest to the origin by solving $\det M(k) = 0$. (Hints: note the structure $M = I + A$ to approximate the determinant; or read *Chaos* **2**, 79 (1992))

- 40.4. **Pinball topological index.** Upgrade your pinball simulator so that it computes the topological index for each orbit it finds.

Chapter 41

Chaotic multiscattering

(A. Wirzba and P. Cvitanović)

WE DISCUSS HERE the semiclassics of scattering in open systems with a finite number of non-overlapping finite scattering regions. Why is this interesting at all? The semiclassics of scattering systems has five advantages compared to the bound-state problems such as the helium quantization discussed in chapter 42.

- For bound-state problem the semiclassical approximation does not respect quantum-mechanical unitarity, and the semiclassical eigenenergies are not real. Here we construct *a manifestly unitary* semiclassical scattering matrix.
- The Weyl-term contributions decouple from the multi-scattering system.
- The close relation to the deterministic escape processes discussed in chapter 1.
- For scattering systems the derivation of cycle expansions is more direct and controlled than in the bound-state case: the semiclassical cycle expansion is the saddle point approximation to the cumulant expansion of the determinant of the exact quantum-mechanical multi-scattering matrix.
- The region of convergence of the semiclassical spectral function is larger than is the case for the bound-state case.

We start by a brief review of the elastic scattering of a point particle from finite collection of non-overlapping scattering regions in terms of the standard textbook scattering theory, and then develop the semiclassical scattering trace formulas and spectral determinants for scattering off N disks in a plane.

41.1 Quantum mechanical scattering matrix

We now specialize to the elastic scattering of a point particle from finite collection of N non-overlapping reflecting disks in a 2-dimensional plane. As the point particle moves freely between the static scatterers, the time-independent Schrödinger equation outside the scattering regions is the Helmholtz equation:

$$(\vec{\nabla}_r^2 + \vec{k}^2)\psi(\vec{r}) = 0, \quad \vec{r} \text{ outside the scattering regions.} \quad (41.1)$$

Here $\psi(\vec{r})$ is the wave function of the point particle at spatial position $\vec{r}$ and $E = \hbar^2 \vec{k}^2 / 2m$ is its energy written in terms of its mass m and the wave vector $\vec{k}$ of the incident wave. For reflecting wall billiards the scattering problem is a boundary value problem with Dirichlet boundary conditions:

$$\psi(\vec{r}) = 0, \quad \vec{r} \text{ on the billiard perimeter} \quad (41.2)$$

As usual for scattering problems, we expand the wave function $\psi(\vec{r})$ in the (2-dimensional) angular momentum eigenfunctions basis

$$\psi(\vec{r}) = \sum_{m=-\infty}^{\infty} \psi_m^k(\vec{r}) e^{-im\Phi_k}, \quad (41.3)$$

where k and Φ_k are the length and angle of the wave vector, respectively. A plane wave in two dimensions expanded in the angular momentum basis is

$$e^{i\vec{k} \cdot \vec{r}} = e^{ikr \cos(\Phi_r - \Phi_k)} = \sum_{m=-\infty}^{\infty} J_m(kr) e^{im(\Phi_r - \Phi_k)}, \quad (41.4)$$

where r and Φ_r denote the distance and angle of the spatial vector $\vec{r}$ as measured in the global 2-dimensional coordinate system.

The m th angular component $J_m(kr) e^{im\Phi_r}$ of a plane wave is split into a superposition of incoming and outgoing 2-dimensional spherical waves by decomposing the ordinary Bessel function $J_m(z)$ into the sum

$$J_m(z) = \frac{1}{2} (H_m^{(1)}(z) + H_m^{(2)}(z)) \quad (41.5)$$

of the Hankel functions $H_m^{(1)}(z)$ and $H_m^{(2)}(z)$ of the first and second kind. For $|z| \gg 1$ the Hankel functions behave asymptotically as:

$$\begin{aligned} H_m^{(2)}(z) &\sim \sqrt{\frac{2}{\pi z}} e^{-i(z - \frac{\pi}{2}m - \frac{\pi}{4})} \quad \text{incoming,} \\ H_m^{(1)}(z) &\sim \sqrt{\frac{2}{\pi z}} e^{+i(z - \frac{\pi}{2}m - \frac{\pi}{4})} \quad \text{outgoing.} \end{aligned} \quad (41.6)$$

Thus for $r \rightarrow \infty$ and k fixed, the m th angular component $J_m(kr)e^{im\Phi_r}$ of the plane wave can be written as superposition of incoming and outgoing 2-dimensional spherical waves:

$$J_m(kr)e^{im\Phi_r} \sim \frac{1}{\sqrt{2\pi kr}} \left[e^{-i(kr - \frac{\pi}{2}m - \frac{\pi}{4})} + e^{i(kr - \frac{\pi}{2}m - \frac{\pi}{4})} \right] e^{im\Phi_r}. \quad (41.7)$$

In terms of the asymptotic (angular momentum) components ψ_m^k of the wave function $\psi(\vec{r})$, the scattering matrix (40.3) is defined as

$$\psi_m^k \sim \frac{1}{\sqrt{2\pi kr}} \sum_{m'=-\infty}^{\infty} \left[\delta_{mm'} e^{-i(kr - \frac{\pi}{2}m' - \frac{\pi}{4})} + S_{mm'} e^{i(kr - \frac{\pi}{2}m' - \frac{\pi}{4})} \right] e^{im'\Phi_r}. \quad (41.8)$$

The matrix element $S_{mm'}$ describes the scattering of an incoming wave with angular momentum m into an outgoing wave with angular momentum m' . If there are no scatterers, then $\mathbf{S} = \mathbf{1}$ and the asymptotic expression of the plane wave $e^{i\vec{k} \cdot \vec{r}}$ in two dimensions is recovered from $\psi(\vec{r})$.

41.1.1 1-disk scattering matrix

In general, $\mathbf{S}$ is nondiagonal and nonseparable. An exception is the 1-disk scatterer. If the origin of the coordinate system is placed at the center of the disk, by (41.5) the m th angular component of the time-independent scattering wave function is a superposition of incoming and outgoing 2-dimensional spherical waves

exercise 40.2

$$\begin{aligned} \psi_m^k &= \frac{1}{2} \left(H_m^{(2)}(kr) + S_{mm} H_m^{(1)}(kr) \right) e^{im\Phi_r} \\ &= \left(J_m(kr) - \frac{i}{2} T_{mm} H_m^{(1)}(kr) \right) e^{im\Phi_r}. \end{aligned}$$

The vanishing (41.2) of the wave function on the disk perimeter

$$0 = J_m(ka) - \frac{i}{2} T_{mm} H_m^{(1)}(ka)$$

yields the 1-disk scattering matrix in analytic form:

$$S_{mm'}^s(k) = \left(1 - \frac{2J_m(ka_s)}{H_m^{(1)}(ka_s)} \right) \delta_{mm'} = -\frac{H_m^{(2)}(ka_s)}{H_m^{(1)}(ka_s)} \delta_{mm'}, \quad (41.9)$$

where $a = a_s$ is radius of the disk and the suffix s indicates that we are dealing with a disk whose label is s . We shall derive a semiclassical approximation to this 1-disk $\mathbf{S}$ -matrix in sect. 41.3.

41.1.2 Multi-scattering matrix

Consider next a scattering region consisting of N non-overlapping disks labeled $s \in \{1, 2, \dots, N\}$, following the notational conventions of sect. 14.6. The strategy is to construct the full $\mathbf{T}$ -matrix (40.3) from the exact 1-disk scattering matrix (41.9) by a succession of coordinate rotations and translations such that at each step the coordinate system is centered at the origin of a disk. Then the $\mathbf{T}$ -matrix in $S_{mm'} = \delta_{mm'} - i T_{mm'}$ can be split into a product over three kinds of matrices,

$$T_{mm'}(k) = \sum_{s,s'=1}^N \sum_{l_s, l_{s'}=-\infty}^{\infty} C_{ml_s}^s(k) \mathbf{M}^{-1}(k)_{l_s l_{s'}}^{ss'} D_{l_{s'} m'}^{s'}(k).$$

The outgoing spherical wave scattered by the disk s is obtained by shifting the global coordinates origin distance R_s to the center of the disk s , and measuring the angle Φ_s with respect to direction $\mathbf{k}$ of the outgoing spherical wave. As in (41.9), the matrix $\mathbf{C}^s$ takes form

$$C_{ml_s}^s = \frac{2i}{\pi a_s} \frac{J_{m-l_s}(kR_s)}{H_{l_s}^{(1)}(ka_s)} e^{im\Phi_s}. \quad (41.10)$$

If we now describe the ingoing spherical wave in the disk s' coordinate frame by the matrix $\mathbf{D}^{s'}$

$$D_{l_{s'} m'}^{s'} = -\pi a_{s'} J_{m'-l_{s'}}(kR_{s'}) J_{l_{s'}}(ka_{s'}) e^{-im'\Phi_{s'}}, \quad (41.11)$$

and apply the Bessel function addition theorem

$$J_m(y+z) = \sum_{\ell=-\infty}^{\infty} J_{m-\ell}(y) J_{\ell}(z),$$

we recover the $\mathbf{T}$ -matrix (41.9) for the single disk $s = s'$, $M = 1$ scattering. The Bessel function sum is a statement of the completeness of the spherical wave basis; as we shift the origin from the disk s to the disk s' by distance $R_{s'}$, we have to reexpand all basis functions in the new coordinate frame.

The labels m and m' refer to the angular momentum quantum numbers of the ingoing and outgoing waves in the global coordinate system, and $l_s, l_{s'}$ refer to the (angular momentum) basis fixed at the s th and s' th scatterer, respectively. Thus, $\mathbf{C}^s$ and $\mathbf{D}^{s'}$ depend on the origin and orientation of the global coordinate system of the 2-dimensional plane as well as on the internal coordinates of the scatterers. As they can be made separable in the scatterer label s , they describe the single scatterer aspects of what, in general, is a multi-scattering problem.

The matrix $\mathbf{M}$ is called the *multi-scattering matrix*. If the scattering problem consists only of one scatterer, $\mathbf{M}$ is simply the unit matrix $M_{l_s l_{s'}}^{ss'} = \delta^{ss'} \delta_{l_s l_{s'}}$. For scattering from more than one scatterer we separate out a “single traversal” matrix $\mathbf{A}$ which transports the scattered wave from a scattering region $\mathcal{M}_s$ to the scattering region $\mathcal{M}_{s'}$,

$$M_{l_s l_{s'}}^{ss'} = \delta^{ss'} \delta_{l_s l_{s'}} - A_{l_s l_{s'}}^{ss'}. \quad (41.12)$$

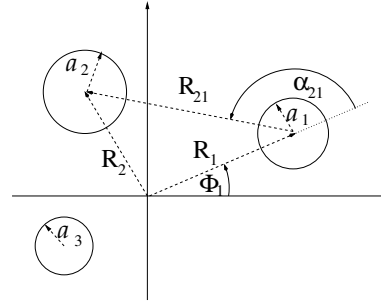


Figure 41.1: Global and local coordinates for a general 3-disk problem.

The matrix $\mathbf{A}^{ss'}$ reads:

$$A_{l_s l_{s'}}^{ss'} = -(1 - \delta^{ss'}) \frac{a_s}{a_{s'}} \frac{J_{l_s}(ka_s)}{H_{l_{s'}}^{(1)}(ka_{s'})} H_{l_s - l_{s'}}^{(1)}(kR_{ss'}) e^{i(l_s \alpha_{s's} - l_{s'}(\alpha_{ss'} - \pi))}. \quad (41.13)$$

Here, a_s is the radius of the s th disk. R_s and Φ_s are the distance and angle, respectively, of the ray from the origin in the 2-dimensional plane to the center of disk s as measured in the global coordinate system. Furthermore, $R_{ss'} = R_{s's}$ is the separation between the centers of the s th and s' th disk and $\alpha_{s's}$ of the ray from the center of disk s to the center of disk s' as measured in the local (body-fixed) coordinate system of disk s (see figure 41.1).

Expanded as a geometrical series about the unit matrix $\mathbf{1}$, the inverse matrix $\mathbf{M}^{-1}$ generates a multi-scattering series in powers of the single-traversal matrix $\mathbf{A}$. All genuine multi-scattering dynamics is contained in the matrix $\mathbf{A}$; by construction $\mathbf{A}$ vanishes for a single-scatterer system.

41.2 N -scatterer spectral determinant

In the following we limit ourselves to a study of the spectral properties of the $\mathbf{S}$ -matrix: resonances, time delays and phase shifts. The resonances are given by the poles of the $\mathbf{S}$ -matrix in the lower complex wave number (k) plane; more precisely, by the poles of the $\mathbf{S}$ on the second Riemann sheet of the complex energy plane. As the $\mathbf{S}$ -matrix is unitary, it is also natural to focus on its total phase shift $\eta(k)$ defined by $\det \mathbf{S} = \exp^{2i\eta(k)}$. The time-delay is proportional to the derivative of the phase shift with respect to the wave number k .

As we are only interested in spectral properties of the scattering problem, it suffices to study $\det \mathbf{S}$. This determinant is basis and coordinate-system independent, whereas the $\mathbf{S}$ -matrix itself depends on the global coordinate system and on the choice of basis for the point particle wave function.

As the $\mathbf{S}$ -matrix is, in general, an infinite dimensional matrix, it is not clear whether the corresponding determinant exists at all. If $\mathbf{T}$ -matrix is trace-class, the determinant does exist. What does this mean?

41.2.1 Trace-class operators

An operator (an infinite-dimensional matrix) is called *trace-class* if and only if, for any choice of orthonormal basis, the sum of the diagonal matrix elements converges absolutely; it is called “Hilbert-Schmidt,” if the sum of the absolute squared diagonal matrix elements converges. Once an operator is diagnosed as trace-class, we are allowed to manipulate it as we manipulate finite-dimensional matrices. We review the theory of trace-class operators in appendix A40; here we will assume that the **T**-matrix (40.3) is trace-class, and draw the conclusions.

If **A** is trace-class, the determinant $\det(\mathbf{1} - z\mathbf{A})$, as defined by the cumulant expansion, exists and is an entire function of z . Furthermore, the determinant is invariant under any unitary transformation.

The cumulant expansion is the analytical continuation (as Taylor expansion in the book-keeping variable z) of the determinant

$$\det(\mathbf{1} - z\mathbf{A}) = \exp[\text{tr} \ln(\mathbf{1} - z\mathbf{A})] = \exp\left(-\sum_{n=1}^{\infty} \frac{z^n}{n} \text{tr}(\mathbf{A}^n)\right).$$

That means

$$\det(\mathbf{1} - z\mathbf{A}) := \sum_{m=0}^{\infty} z^m Q_m(\mathbf{A}), \quad (41.14)$$

where the cumulants $Q_m(\mathbf{A})$ satisfy the Plemelj-Smithies recursion formula (A40.22), a generalization of Newton’s formula to determinants of infinite-dimensional matrices,

$$\begin{aligned} Q_0(\mathbf{A}) &= 1 \\ Q_m(\mathbf{A}) &= -\frac{1}{m} \sum_{j=1}^m Q_{m-j}(\mathbf{A}) \text{tr}(\mathbf{A}^j) \quad \text{for } m \geq 1, \end{aligned} \quad (41.15)$$

in terms of cumulants of order $n < m$ and traces of order $n \leq m$. Because of the trace-class property of **A**, all cumulants and traces exist separately.

For the general case of $N < \infty$ non-overlapping scatterers, the **T**-matrix can be shown to be trace-class, so the determinant of the **S**-matrix is well defined. What does trace-class property mean for the corresponding matrices **C**^s, **D**^s and **A**^{ss′}? Manipulating the operators as though they were finite matrices, we can perform the following transformations:

$$\begin{aligned} \det \mathbf{S} &= \det(\mathbf{1} - i\mathbf{C}\mathbf{M}^{-1}\mathbf{D}) \\ &= \text{Det}(\mathbf{1} - i\mathbf{M}^{-1}\mathbf{D}\mathbf{C}) = \text{Det}(\mathbf{M}^{-1}(\mathbf{M} - i\mathbf{D}\mathbf{C})) \\ &= \frac{\text{Det}(\mathbf{M} - i\mathbf{D}\mathbf{C})}{\text{Det}(\mathbf{M})} \dots \end{aligned} \quad (41.16)$$

In the first line of (41.16) the determinant is taken over small ℓ (the angular momentum with respect to the global system). In the remainder of (41.16) the determinant is evaluated over the multiple indices $L_s = (s, l_s)$. In order to signal this difference we use the following notation: $\det \dots$ and $\text{tr} \dots$ refer to the $|\ell\rangle$ space, $\text{Det} \dots$ and $\text{Tr} \dots$ refer to the multiple index space. The matrices in the multiple index space are expanded in the complete basis $\{|L_s\rangle\} = \{|s, \ell_s\rangle\}$ which refers for fixed index s to the origin of the s th scatterer and not any longer to the origin of the 2-dimensional plane.

Let us explicitly extract the product of the determinants of the subsystems from the determinant of the total system (41.16):

$$\begin{aligned} \det \mathbf{S} &= \frac{\text{Det}(\mathbf{M} - i\mathbf{D}\mathbf{C})}{\text{Det}(\mathbf{M})} \\ &= \frac{\text{Det}(\mathbf{M} - i\mathbf{D}\mathbf{C})}{\text{Det} \mathbf{M}} \frac{\prod_{s=1}^N \det \mathbf{S}^s}{\prod_{s=1}^N \det \mathbf{S}^s} \\ &= \left(\prod_{s=1}^N \det \mathbf{S}^s \right) \frac{\text{Det}(\mathbf{M} - i\mathbf{D}\mathbf{C}) / \prod_{s=1}^N \det \mathbf{S}^s}{\text{Det} \mathbf{M}}. \end{aligned} \quad (41.17)$$

The final step in the reformulation of the determinant of the $\mathbf{S}$ -matrix of the N -scatterer problem follows from the unitarity of the $\mathbf{S}$ -matrix. The unitarity of $\mathbf{S}^\dagger(k^*)$ implies for the determinant

$$\det(\mathbf{S}(k^*)^\dagger) = 1/\det \mathbf{S}(k), \quad (41.18)$$

where this manipulation is allowed because the $\mathbf{T}$ -matrix is trace-class. The unitarity condition should apply for the $\mathbf{S}$ -matrix of the total system, $\mathbf{S}$, as for the each of the single subsystems, $\mathbf{S}^s$, $s = 1, \dots, N$. In terms of the result of (41.17), this implies

$$\frac{\text{Det}(\mathbf{M}(k) - i\mathbf{D}(k)\mathbf{C}(k))}{\prod_{s=1}^N \det \mathbf{S}^s} = \text{Det}(\mathbf{M}(k^*)^\dagger)$$

since all determinants in (41.17) exist separately and since the determinants $\det \mathbf{S}^s$ respect unitarity by themselves. Thus, we finally have

$$\det \mathbf{S}(k) = \left\{ \prod_{s=1}^N (\det \mathbf{S}^s(k)) \right\} \frac{\text{Det} \mathbf{M}(k^*)^\dagger}{\text{Det} \mathbf{M}(k)}, \quad (41.19)$$

where all determinants exist separately.

In summary: We assumed a scattering system of a *finite* number of *non-overlapping* scatterers which can be of different shape and size, but are all of finite extent. We assumed the trace-class character of the $\mathbf{T}$ -matrix belonging to the total system and of the single-traversal matrix $\mathbf{A}$ and finally unitarity of the $\mathbf{S}$ -matrices of the complete and all subsystems.

What can one say about the point-particle scattering from a finite number of scatterers of arbitrary shape and size? As long as each of $N < \infty$ single scatterers

has a finite spatial extent, i.e., can be covered by a finite disk, the total system has a finite spatial extent as well. Therefore, it too can be put inside a circular domain of finite radius b , e.g., inside a single disk. If the impact parameter of the point particle measured with respect to the origin of this disk is larger than the disk size (actually larger than $(e/2) \times b$), then the $\mathbf{T}$ matrix elements of the N -scatterer problem become very small. If the wave number k is kept fixed, the modulus of the *diagonal* matrix elements, $|T_{mm}|$ with the angular momentum $m > (e/2)kb$, is bounded by the corresponding quantity of the covering disk.

41.2.2 Quantum cycle expansions

In formula (41.19) the genuine multi-scattering terms are separated from the single-scattering ones. We focus on the multi-scattering terms, i.e., on the ratio of the determinants of the multi-scattering matrix $\mathbf{M} = \mathbf{1} - \mathbf{A}$ in (41.19), since they are the origin of the periodic orbit sums in the semiclassical reduction. The resonances of the multi-scattering system are given by the zeros of $\text{Det } \mathbf{M}(k)$ in the lower complex wave number plane.

In order to set up the problem for the semiclassical reduction, we express the determinant of the multi-scattering matrix in terms of the traces of the powers of the matrix $\mathbf{A}$, by means of the cumulant expansion (41.14). Because of the finite number $N \geq 2$ of scatterers $\text{Tr}(\mathbf{A}^n)$ receives contributions corresponding to all periodic itineraries $s_1 s_2 s_3 \cdots s_{n-1} s_n$ of total symbol length n with an alphabet $s_i \in \{1, 2, \dots, N\}$, of N symbols,

$$\begin{aligned} \text{tr } \mathbf{A}^{s_1 s_2} \mathbf{A}^{s_2 s_3} \cdots \mathbf{A}^{s_{n-1} s_n} \mathbf{A}^{s_n s_1} \\ = \sum_{l_{s_1}=-\infty}^{+\infty} \sum_{l_{s_2}=-\infty}^{+\infty} \cdots \sum_{l_{s_n}=-\infty}^{+\infty} A_{l_{s_1} l_{s_2}}^{s_1 s_2} A_{l_{s_2} l_{s_3}}^{s_2 s_3} \cdots A_{l_{s_{n-1}} l_{s_n}}^{s_{n-1} s_n} A_{l_{s_n} l_{s_1}}^{s_n s_1}. \end{aligned} \quad (41.20)$$

Remember our notation that the trace $\text{tr}(\cdots)$ refers only to the $|l\rangle$ space. By construction $\mathbf{A}$ describes only scatterer-to-scatterer transitions, so the symbolic dynamics has to respect the no-self-reflection pruning rule: for admissible itineraries the successive symbols have to be different. This rule is implemented by the factor $1 - \delta^{s s'}$ in (41.13).

The trace $\text{Tr } \mathbf{A}^n$ is the sum of all itineraries of length n ,

$$\text{Tr } \mathbf{A}^n = \sum_{\{s_1 s_2 \cdots s_n\}} \text{tr } \mathbf{A}^{s_1 s_2} \mathbf{A}^{s_2 s_3} \cdots \mathbf{A}^{s_{n-1} s_n} \mathbf{A}^{s_n s_1}. \quad (41.21)$$

We will show for the N -disk problem that these periodic itineraries correspond in the semiclassical limit, $ka_{s_i} \gg 1$, to *geometrical* periodic orbits with the same symbolic dynamics.

For periodic orbits with creeping sections the symbolic alphabet has to be extended, see sect. 41.3.1. Furthermore, depending on the geometry, there might be nontrivial pruning rules based on the so called ghost orbits, see sect. 41.4.1.

41.2.3 Symmetry reductions

The determinants over the multi-scattering matrices run over the multiple index L of the multiple index space. This is the proper form for the symmetry reduction (in the multiple index space), e.g., if the scatterer configuration is characterized by a discrete symmetry group G , we have

$$\text{Det } \mathbf{M} = \prod_{\alpha} (\det \mathbf{M}_{D_{\alpha}}(k))^{d_{\alpha}},$$

where the index α runs over all conjugate classes of the symmetry group G and D_{α} is the α th representation of dimension d_{α} . The symmetry reduction on the exact quantum mechanical level is the same as for the deterministic evolution operators spectral determinant factorization (25.22) of sect. 25.4.2.

41.3 Semiclassical 1-disk scattering

We start by focusing on the single-scatterer problem. In order to be concrete, we will consider the semiclassical reduction of the scattering of a single disk in plane.

Instead of calculating the semiclassical approximation to the determinant of the one-disk system scattering matrix (41.9), we do so for

$$\mathbf{d}(k) \equiv \frac{1}{2\pi i} \frac{d}{dk} \ln \det \mathbf{S}^1(ka) = \frac{1}{2\pi i} \frac{d}{dk} \text{tr} \left(\ln \mathbf{S}^1(ka) \right) \quad (41.22)$$

the so called *time delay*.

$$\begin{aligned} \mathbf{d}(k) &= \frac{1}{2\pi i} \frac{d}{dk} \text{tr} \left(\ln \det \mathbf{S}^1(ka) \right) = \frac{1}{2\pi i} \sum_m \left(\frac{H_m^{(1)}(ka)}{H_m^{(2)}(ka)} \frac{d}{dk} \frac{H_m^{(2)}(ka)}{H_m^{(1)}(ka)} \right) \\ &= \frac{a}{2\pi i} \sum_m \left(\frac{H_m^{(2)'}(ka)}{H_m^{(2)}(ka)} - \frac{H_m^{(1)'}(ka)}{H_m^{(1)}(ka)} \right). \end{aligned} \quad (41.23)$$

Here the prime denotes the derivative with respect to the argument of the Hankel functions. Let us introduce the abbreviation

$$\chi_v = \frac{H_v^{(2)'}(ka)}{H_v^{(2)}(ka)} - \frac{H_v^{(1)'}(ka)}{H_v^{(1)}(ka)}. \quad (41.24)$$

We apply the Watson contour method to (41.23)

$$\mathbf{d}(k) = \frac{a_j}{2\pi i} \sum_{m=-\infty}^{+\infty} \chi_m = \frac{a_j}{2\pi i} \frac{1}{2i} \oint_C dv \frac{e^{-iv\pi}}{\sin(v\pi)} \chi_v. \quad (41.25)$$

Here the contour C encircles in a counter-clockwise manner a small semiinfinite strip D which completely covers the real v -axis but which only has a small finite extent into the positive and negative imaginary v direction. The contour C is then split up in the path above and below the real v -axis such that

$$\mathbf{d}(k) = \frac{a}{4\pi i} \left\{ \int_{-\infty+i\epsilon}^{+\infty+i\epsilon} dv \frac{e^{-iv\pi}}{\sin(v\pi)} \chi_v - \int_{-\infty-i\epsilon}^{+\infty-i\epsilon} dv \frac{e^{-iv\pi}}{\sin(v\pi)} \chi_v \right\}.$$

Then, we perform the substitution $v \rightarrow -v$ in the second integral so as to get

$$\begin{aligned} \mathbf{d}(k) &= \frac{a}{4\pi} \left\{ \int_{-\infty+i\epsilon}^{+\infty+i\epsilon} dv \frac{e^{-iv\pi}}{\sin(v\pi)} \chi_v + \int_{-\infty-i\epsilon}^{+\infty-i\epsilon} dv \frac{e^{+iv\pi}}{\sin(v\pi)} \chi_{-v} \right\} \\ &= \frac{a}{2\pi i} \left\{ 2 \int_{-\infty+i\epsilon}^{+\infty+i\epsilon} dv \frac{e^{2iv\pi}}{1 - e^{2iv\pi}} \chi_v + \int_{-\infty}^{+\infty} dv \chi_v \right\}, \end{aligned} \quad (41.26)$$

where we used the fact that $\chi_{-v} = \chi_v$. The contour in the last integral can be deformed to pass over the real v -axis since its integrand has no Watson denominator.

We will now approximate the last expression semiclassically, i.e., under the assumption $ka \gg 1$. As the two contributions in the last line of (41.26) differ by the presence or absence of the Watson denominator, they will have to be handled semiclassically in different ways: the first will be closed in the upper complex plane and evaluated at the poles of χ_v , the second integral will be evaluated on the real v -axis under the Debye approximation for Hankel functions.

We will now work out the first term. The poles of χ_v in the upper complex plane are given by the zeros of $H_v^{(1)}(ka)$ which will be denoted by $v_\ell(ka)$ and by the zeros of $H_v^{(2)}(ka)$ which we will denote by $-\bar{v}_\ell(ka)$, $\ell = 1, 2, 3, \dots$. In the Airy approximation to the Hankel functions they are given by

$$v_\ell(ka) = ka + i\alpha_\ell(ka), \quad (41.27)$$

$$-\bar{v}_\ell(ka) = -ka + i(\alpha_\ell(k^*a))^* = -(v_\ell(k^*a))^*, \quad (41.28)$$

with

$$\begin{aligned} i\alpha_\ell(ka) &= e^{i\frac{\pi}{3}} \left(\frac{ka}{6} \right)^{1/3} q_\ell - e^{-i\frac{\pi}{3}} \left(\frac{6}{ka} \right)^{1/3} \frac{q_\ell^2}{180} - \frac{1}{70ka} \left(1 - \frac{q_\ell^3}{30} \right) \\ &+ e^{i\frac{\pi}{3}} \left(\frac{6}{ka} \right)^{\frac{5}{3}} \frac{1}{3150} \left(\frac{29q_\ell}{6^2} - \frac{281q_\ell^4}{180 \cdot 6^3} \right) + \dots \end{aligned} \quad (41.29)$$

Here q_ℓ labels the zeros of the Airy integral

$$A(q) \equiv \int_0^\infty d\tau \cos(q\tau - \tau^3) = 3^{-1/3} \pi \text{Ai}(-3^{-1/3} q), \quad (41.30)$$

with $\text{Ai}(z)$ being the standard Airy function; approximately, $q_\ell \approx 6^{1/3} [3\pi(\ell - 1/4)]^{2/3}/2$. In order to keep the notation simple, we will abbreviate $\nu_\ell \equiv \nu_\ell(ka)$ and $\bar{\nu}_\ell \equiv \bar{\nu}_\ell(ka)$. Thus the first term of (41.26) becomes finally

$$\frac{a}{2\pi i} \left\{ 2 \int_{-\infty+i\epsilon}^{+\infty+i\epsilon} d\nu \frac{e^{2i\nu\pi}}{1 - e^{2i\nu\pi}} \chi_\nu \right\} = 2a \sum_{\ell=1}^{\infty} \left(\frac{e^{2i\nu_\ell\pi}}{1 - e^{2i\nu_\ell\pi}} + \frac{e^{-2i\bar{\nu}_\ell\pi}}{1 - e^{-2i\bar{\nu}_\ell\pi}} \right).$$

In the second term of (41.26) we will insert the Debye approximations for the Hankel functions:

$$H_\nu^{(1/2)}(x) \sim \sqrt{\frac{2}{\pi \sqrt{x^2 - \nu^2}}} \exp\left(\pm i \sqrt{x^2 - \nu^2} \mp i\nu \arccos \frac{\nu}{x} \mp i\frac{\pi}{4}\right) \quad \text{for } |x| > \nu \quad (41.31)$$

$$H_\nu^{(1/2)}(x) \sim \mp i \sqrt{\frac{2}{\pi \sqrt{\nu^2 - x^2}}} \exp\left(-\sqrt{\nu^2 - x^2} + \nu \text{ArcCosh} \frac{\nu}{x}\right) \quad \text{for } |x| < \nu.$$

Note that for $\nu > ka$ the contributions in χ_ν cancel. Thus the second integral of (41.26) becomes

$$\begin{aligned} \frac{a}{2\pi i} \int_{-\infty}^{+\infty} d\nu \chi_\nu &= \frac{a}{2\pi i} \int_{-ka}^{+ka} d\nu \frac{(-2i)}{a} \frac{d}{dk} \left(\sqrt{k^2 a^2 - \nu^2} - \nu \arccos \frac{\nu}{ka} \right) + \dots \\ &= -\frac{1}{k\pi} \int_{-ka}^{+ka} d\nu \sqrt{k^2 a^2 - \nu^2} + \dots = -\frac{a^2}{2} k + \dots, \end{aligned} \quad (41.32)$$

where $\dots$ takes care of the polynomial corrections in the Debye approximation and the boundary correction terms in the ν integration.

In summary, the semiclassical approximation to $\mathbf{d}(k)$ reads

$$\mathbf{d}(k) = 2a \sum_{\ell=1}^{\infty} \left(\frac{e^{2i\nu_\ell\pi}}{1 - e^{2i\nu_\ell\pi}} + \frac{e^{-2i\bar{\nu}_\ell\pi}}{1 - e^{-2i\bar{\nu}_\ell\pi}} \right) - \frac{a^2}{2} k + \dots.$$

Using the definition of the time delay (41.22), we get the following expression for $\det \mathbf{S}^1(ka)$:

$$\begin{aligned} \ln \det \mathbf{S}^1(ka) - \lim_{k_0 \rightarrow 0} \ln \det \mathbf{S}^1(k_0 a) &= 2\pi i a \int_0^k d\tilde{k} \left(-\frac{a\tilde{k}}{2} + 2 \sum_{\ell=1}^{\infty} \left(\frac{e^{i2\pi\nu_\ell(\tilde{k}a)}}{1 - e^{i2\pi\nu_\ell(\tilde{k}a)}} + \frac{e^{-i2\pi\bar{\nu}_\ell(\tilde{k}a)}}{1 - e^{-i2\pi\bar{\nu}_\ell(\tilde{k}a)}} \right) \right) + \dots \\ &\sim -2\pi i N(k) + 2 \sum_{\ell=1}^{\infty} \int_0^k d\tilde{k} \frac{d}{d\tilde{k}} \left\{ -\ln(1 - e^{i2\pi\nu_\ell(\tilde{k}a)}) + \ln(1 - e^{-i2\pi\bar{\nu}_\ell(\tilde{k}a)}) \right\} + \dots, \end{aligned} \quad (41.33)$$

where in the last expression it has been used that semiclassically $\frac{d}{dk}v_\ell(ka) \sim \frac{d}{dk}\bar{v}_\ell(ka) \sim a$ and that the Weyl term for a single disk of radius a goes like $N(k) = \pi a^2 k^2 / (4\pi) + \dots$ (the next terms come from the boundary terms in the v -integration in (41.32)). Note that for the lower limit, $k_0 \rightarrow 0$, we have two simplifications: First,

$$\begin{aligned} \lim_{k_0 \rightarrow 0} S_{mm'}^1(k_0 a) &= \lim_{k_0 \rightarrow 0} \frac{-H_m^{(2)}(k_0 a)}{H_m^{(1)}(k_0 a)} \delta_{mm'} = 1 \times \delta_{mm'} \quad \forall m, m' \\ &\rightarrow \lim_{k_0 \rightarrow 0} \det \mathbf{S}^1(k_0 a) = 1. \end{aligned}$$

Secondly, for $k_0 \rightarrow 0$, the two terms in the curly bracket of (41.33) cancel.

41.3.1 1-disk spectrum interpreted; pure creeping

To summarize: the semiclassical approximation to the determinant $\mathbf{S}^1(ka)$ is given by

$$\det \mathbf{S}^1(ka) \sim e^{-i2\pi N(k)} \frac{\prod_{\ell=1}^{\infty} (1 - e^{-2i\pi \bar{v}_\ell(ka)})^2}{\prod_{\ell=1}^{\infty} (1 - e^{2i\pi v_\ell(ka)})^2}, \quad (41.34)$$

with

$$\begin{aligned} v_\ell(ka) &= ka + i\alpha_\ell(ka) = ka + e^{+i\pi/3}(ka/6)^{1/3}q_\ell + \dots \\ \bar{v}_\ell(ka) &= ka - i(\alpha_\ell(k^*a))^* = ka + e^{-i\pi/3}(ka/6)^{1/3}q_\ell + \dots \\ &= (v_\ell(k^*a))^* \end{aligned}$$

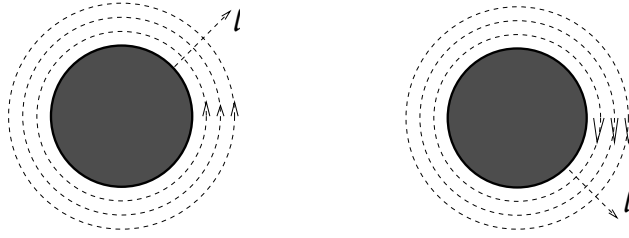
and $N(ka) = (\pi a^2 k^2)/4\pi + \dots$ the leading term in the Weyl approximation for the staircase function of the wavenumber eigenvalues in the disk interior. From the point of view of the scattering particle, the interior domains of the disks are excluded relatively to the free evolution without scattering obstacles. Therefore the negative sign in front of the Weyl term. For the same reason, the subleading boundary term has here a Neumann structure, although the disks have Dirichlet boundary conditions.

Let us abbreviate the r.h.s. of (41.34) for a disk s as

$$\det \mathbf{S}^s(ka_s) \sim \left(e^{-i\pi N(ka_s)} \right)^2 \frac{\tilde{Z}_\ell^s(k^*a_s)^*}{\tilde{Z}_\ell^s(ka_s)} \frac{\tilde{Z}_r^s(k^*a_s)^*}{\tilde{Z}_r^s(ka_s)}, \quad (41.35)$$

where $\tilde{Z}_\ell^s(ka_s)$ and $\tilde{Z}_r^s(ka_s)$ are the *diffractive* zeta functions (here and in the following we will label semiclassical zeta functions *with* diffractive corrections by a tilde) for creeping orbits around the s th disk in the left-handed sense and the right-handed sense, respectively (see figure 41.2). The two orientations of the creeping orbits are the reason for the exponents 2 in (41.34). Equation (41.34) describes

Figure 41.2: Right- and left-handed diffractive creeping paths of increasing mode number ℓ for a single disk.



the semiclassical approximation to the incoherent part (= the curly bracket on the r.h.s.) of the exact expression (41.19) for the case that the scatterers are disks.

In the following we will discuss the semiclassical resonances in the 1-disk scattering problem with Dirichlet boundary conditions, i.e. the so-called shape resonances. The quantum mechanical resonances are the poles of the S -matrix in the complex k -plane. As the 1-disk scattering problem is separable, the S -matrix is already diagonalized in the angular momentum eigenbasis and takes the simple form (41.9). The exact quantummechanical poles of the scattering matrix are therefore given by the zeros k_{nm}^{res} of the Hankel functions $H_m^{(1)}(ka)$ in the lower complex k plane which can be labeled by two indices, m and n , where m denotes the angular quantum number of the Hankel function and n is a radial quantum number. As the Hankel functions have to vanish at specific k values, one cannot use the usual Debye approximation as semiclassical approximation for the Hankel function, since this approximation only works in case the Hankel function is dominated by only one saddle. However, for the vanishing of the Hankel function, one has to have the interplay of two saddles, thus an Airy approximation is needed as in the case of the creeping poles discussed above. The Airy approximation of the Hankel function $H_\nu^{(1)}(ka)$ of complex-valued index ν reads

$$H_\nu^{(1)}(ka) \sim \frac{2}{\pi} e^{-i\frac{\pi}{3}} \left(\frac{6}{ka} \right)^{1/3} A(q^{(1)}),$$

with

$$q^{(1)} = e^{-i\frac{\pi}{3}} \left(\frac{6}{ka} \right)^{1/3} (\nu - ka) + O((ka)^{-1}).$$

Hence the zeros ν_ℓ of the Hankel function in the complex ν plane follow from the zeros q_ℓ of the Airy integral $A(q)$ (see (41.30)). Thus if we set $\nu_\ell = m$ (with m integer), we have the following semiclassical condition on k^{res}

$$\begin{aligned} m &\sim k^{\text{res}} a + i\alpha_\ell (k^{\text{res}} a) \\ &= e^{i\frac{\pi}{3}} \left(\frac{k^{\text{res}} a}{6} \right)^{1/3} q_\ell - e^{-i\frac{\pi}{3}} \left(\frac{6}{k^{\text{res}} a} \right)^{1/3} \frac{q_\ell^2}{180} - \frac{1}{70k^{\text{res}} a} \left(1 - \frac{q_\ell^3}{30} \right) \\ &\quad + e^{i\frac{\pi}{3}} \left(\frac{6}{k^{\text{res}} a} \right)^{5/3} \frac{1}{3150} \left(\frac{29q_\ell}{6^2} - \frac{281q_\ell^4}{180 \cdot 6^3} \right) + \dots, \end{aligned}$$

with $l = 1, 2, 3, \dots$.

(41.36)

For a given index l this is equivalent to

$$0 \sim 1 - e^{(ik^{\text{res}} - \alpha_\ell)2\pi a},$$

Figure 41.3: The shape resonances of the 1-disk system in the complex k plane in units of the disk radius a . The boxes label the exact quantum mechanical resonances (given by the zeros of $H_m^{(1)}(ka)$ for $m = 0, 1, 2$), the crosses label the diffractive semiclassical resonances (given by the zeros of the creeping formula in the Airy approximation (41.36) up to the order $O([ka]^{1/3})$).

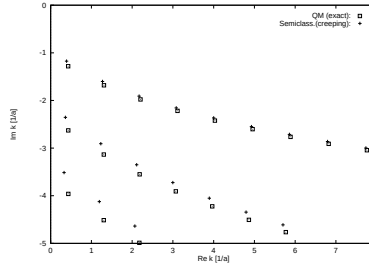
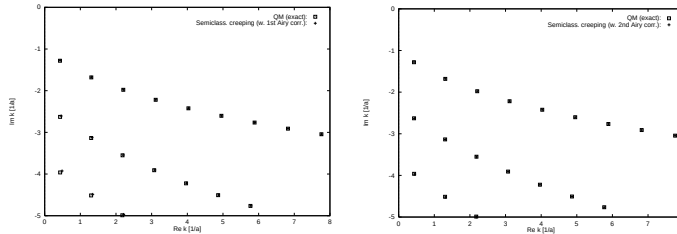


Figure 41.4: Same as in figure 41.3. However, the subleading terms in the Airy approximation (41.36) are taken into account up to the order $O([ka]^{-1/3})$ (upper panel) and up to order $O([ka]^{-1})$ (lower panel).



the de-Broglie condition on the wave function that encircles the disk. Thus the semiclassical resonances of the 1-disk problem are given by the zeros of the following product

$$\prod_{l=1}^{\infty} (1 - e^{(ik - \alpha_l)2\pi a}),$$

which is of course nothing else than $\tilde{Z}_{1\text{-disk}}(k)$, the semiclassical diffraction zeta function of the 1-disk scattering problem, see (41.35). Note that this expression includes just the pure creeping contribution and no genuine geometrical parts. Because of

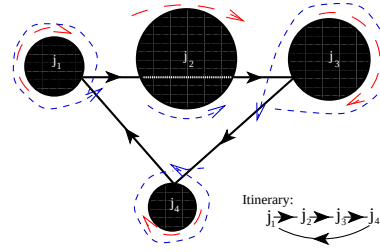
$$H_{-m}^{(1)}(ka) = (-1)^m H_m^{(1)}(ka),$$

the zeros are doubly degenerate if $m \neq 0$, corresponding to right- and left handed creeping turns. The case $m = 0$ is unphysical, since all zeros of the Hankel function $H_0^{(1)}(ka)$ have negative real value.

From figure 41.3 one notes that the creeping terms in the Airy order $O([ka]^{1/3})$, which are used in the Keller construction, systematically underestimate the magnitude of the imaginary parts of the exact data. However, the creeping data become better for increasing $\text{Re } k$ and decreasing $|\text{Im } k|$, as they should as semiclassical approximations.

In the upper panel of figure 41.4 one sees the change, when the next order in the Airy approximation (41.36) is taken into account. The approximation is nearly perfect, especially for the leading row of resonances. The second Airy approximation using (41.36) up to order $O([ka]^{-1})$ is perfect up to the drawing scale of figure 41.4 (lower panel).

Figure 41.5: A 4-disk problem with three specular reflections, one ghost tunneling, and distinct creeping segments from which all associated creeping paths can be constructed.



41.4 From quantum cycle to semiclassical cycle

The procedure for the semiclassical approximation of a general periodic itinerary (41.20) of period n is somewhat laborious, and we will only sketch the procedure here. It follows, in fact, rather closely the methods developed for the semiclassical reduction of the determinant of the 1-disk system.

The quantum cycle

$$\text{tr } \mathbf{A}^{s_1 s_2} \cdots \mathbf{A}^{s_m s_1} = \sum_{l_{s_1}=-\infty}^{\infty} \cdots \sum_{l_{s_m}=-\infty}^{\infty} A_{l_{s_1} l_{s_2}}^{s_1 s_2} \cdots A_{l_{s_m} l_{s_1}}^{s_m s_1}$$

still has the structure of a “multi-trace” with respect to angular momentum.

Each of the sums $\sum_{l_{s_i}=-\infty}^{\infty}$ – as in the 1-disk case – is replaced by a *Watson contour* resummation in terms of complex angular momentum ν_{s_i} . Then the paths below the real ν_{s_i} -axes are transformed to paths above these axes, and the integrals split into expressions *with* and *without* an explicit Watson $\sin(\nu_{s_i}\pi)$ denominator.

1. In the $\sin(\nu_{s_i}\pi)$ -independent integrals we replace all Hankel and Bessel functions by Debye approximations. Then we evaluate the expression in the saddle point approximation: either left or right *specular reflection* at disk s_i or *ghost tunneling* through disk s_i result.
2. For the $\sin(\nu_{s_i}\pi)$ -dependent integrals, we close the contour in the upper ν_{s_i} plane and evaluate the integral at the residua $H_{\nu_{s_i}}^{(1)}(ka_{s_i})=0$. Then we use the Airy approximation for $J_{\nu_{s_i}}(ka_{s_i})$ and $H_{\nu_{s_i}}^{(1)}(ka_{s_i})$: left and right *creeping paths* around disk s_i result.

In the above we have assumed that no grazing geometrical paths appear. If they do show up, the analysis has to be extended to the case of coinciding saddles between the geometrical paths with $\pi/2$ angle reflection from the disk surface and paths with direct ghost tunneling through the disk.

There are three possibilities of “semiclassical” contact of the point particle with the disk s_i :

1. either geometrical which in turn splits into three alternatives

- (a) *specular reflection* to the right,
 - (b) *specular reflection* to the left,
 - (c) or ‘*ghost tunneling*’ where the latter induce the nontrivial pruning rules (as discussed above)
2. or *right-handed creeping turns*
 3. or *left-handed creeping turns*,

see figure 41.5. The specular reflection to the right is linked to left-handed creeping paths with at least one knot. The specular reflection to the left matches a right-handed creeping paths with at least one knot, whereas the shortest left- and right-handed creeping paths in the ghost tunneling case are topologically trivial. In fact, the topology of the creeping paths encodes the choice between the three alternatives for the geometrical contact with the disk. This is the case for the simple reason that creeping sections have to be positive definite in length: the creeping amplitude has to decrease during the creeping process, as tangential rays are constantly emitted. In mathematical terms, it means that the creeping angle has to be positive. Thus, the positivity of the *two* creeping angles for the shortest left *and* right turn uniquely specifies the topology of the creeping sections which in turn specifies which of the three alternatives, either specular reflection to the right or to the left or straight “ghost” tunneling through disk j , is realized for the semiclassical geometrical path. Hence, the existence of a unique saddle point is guaranteed.

In order to be concrete, we will restrict ourselves in the following to the scattering from $N < \infty$ non-overlapping *disks* fixed in the 2-dimensional plane. The semiclassical approximation of the periodic itinerary

$$\text{tr } \mathbf{A}^{s_1 s_2} \mathbf{A}^{s_2 s_3} \dots \mathbf{A}^{s_{n-1} s_n} \mathbf{A}^{s_n s_1}$$

becomes a standard periodic orbit labeled by the symbol sequence $s_1 s_2 \dots s_n$. Depending on the geometry, the individual legs $s_{i-1} \rightarrow s_i \rightarrow s_{i+1}$ result either from a standard specular reflection at disk s_i or from a ghost path passing straight through disk s_i . If furthermore creeping contributions are taken into account, the symbolic dynamics has to be generalized from single-letter symbols $\{s_i\}$ to triple-letter symbols $\{s_i, \sigma_i \times \ell_i\}$ with $\ell_i \geq 1$ integer valued and $\sigma_i = 0, \pm 1$ ¹ By definition, the value $\sigma_i = 0$ represents the non-creeping case, such that $\{s_i, 0 \times \ell_i\} = \{s_i, 0\} = \{s_i\}$ reduces to the old single-letter symbol. The magnitude of a nonzero ℓ_i corresponds to creeping sections of mode number $|\ell_i|$, whereas the sign $\sigma_i = \pm 1$ signals whether the creeping path turns around the disk s_i in the positive or negative sense. Additional full creeping turns around a disk s_i can be summed up as a geometrical series; therefore they do not lead to the introduction of a further symbol.

41.4.1 Ghost contributions

An itinerary with a semiclassical ghost section at, say, disk s_i can be shown to have the same weight as the corresponding itinerary without the s_i th symbol.

¹Actually, these are double-letter symbols as σ_i and ℓ_i are only counted as a product.

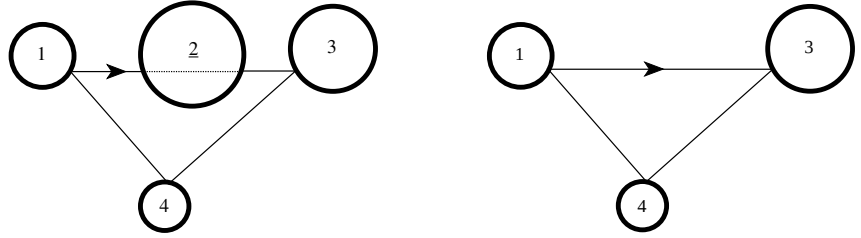


Figure 41.6: (a) The ghost itinerary (1, 2, 3, 4). (b) The parent itinerary (1, 3, 4).

Thus, semiclassically, they cancel each other in the $\text{Tr} \ln(\mathbf{1} - \mathbf{A})$ expansion, where they are multiplied by the permutation factor n/r with the integer r counting the repeats. For example, let (1, 2, 3, 4) be a non-repeated periodic itinerary with a ghost section at disk 2 stemming from the 4th-order trace $\text{Tr} A^4$. By convention, an underlined disk index signals a ghost passage (as in figure 41.6a), with corresponding semiclassical ghost traversal matrices also underlined, $\underline{A}^{i,i+1} \underline{A}^{i+1,i+2}$. Then its semiclassical, geometrical contribution to $\text{Tr} \ln(\mathbf{1} - \mathbf{A})$ cancels exactly against the one of its “parent” itinerary (1, 3, 4) (see figure 41.6b) resulting from the 3rd-order trace:

$$\begin{aligned}
 & -\frac{1}{4} \left(4 \underline{A}^{1,2} \underline{A}^{2,3} A^{3,4} A^{4,1} \right) - \frac{1}{3} \left(3 A^{1,3} A^{3,4} A^{4,1} \right) \\
 & = (+1 - 1) A^{1,3} A^{3,4} A^{4,1} = 0.
 \end{aligned}$$

The prefactors $-1/3$ and $-1/4$ are due to the expansion of the logarithm, the factors 3 and 4 inside the brackets result from the cyclic permutation of the periodic itineraries, and the cancellation stems from the rule

$$\dots \underline{A}^{i,i+1} \underline{A}^{i+1,i+2} \dots = \dots \left(-A^{i,i+2} \right) \dots \quad (41.37)$$

The reader might study more complicated examples and convince herself that the rule (41.37) is sufficient to cancel any primary or repeated periodic orbit with one or more ghost sections completely out of the expansion of $\text{Tr} \ln(\mathbf{1} - \mathbf{A})$ and therefore also out of the cumulant expansion in the semiclassical limit: Any periodic orbit of period m with $n(< m)$ ghost sections is cancelled by the sum of all ‘parent’ periodic orbits of period $m - i$ (with $1 \leq i \leq n$ and i ghost sections removed) weighted by their cyclic permutation factor and by the prefactor resulting from the *trace-log* expansion. This is the way in which the nontrivial pruning for the N -disk billiards can be derived from the exact quantum mechanical expressions in the semiclassical limit. Note that there must exist at least one index i in any given *periodic* itinerary which corresponds to a non-ghost section, since otherwise the itinerary in the semiclassical limit could only be straight and therefore nonperiodic. Furthermore, the series in the ghost cancelation has to stop at the 2nd-order trace, $\text{Tr} A^2$, as $\text{Tr} A$ itself vanishes identically in the full domain which is considered here.

41.5 Heisenberg uncertainty

Where is the boundary $ka \approx 2^{m-1} \bar{L}/a$ coming from?

This boundary follows from a combination of the uncertainty principle with ray optics and the non-vanishing value for the topological entropy of the 3-disk repeller. When the wave number k is fixed, quantum mechanics can only resolve the deterministic repelling set up to the critical topological order n . The quantum wave packet which explores the repelling set has to disentangle 2^n different sections of size $d \sim a/2^n$ on the “visible” part of the disk surface (which is of order a) between any two successive disk collisions. Successive collisions are separated spatially by the mean flight length $\bar{L}$, and the flux spreads with a factor $\bar{L}/a$. In other words, the uncertainty principle bounds the maximal sensible truncation in the cycle expansion order by the highest quantum resolution attainable for a given wavenumber k .

Commentary

Remark 41.1. Sources. This chapter is based in its entirety on ref. [21]; the reader is referred to the full exposition for the proofs and discussion of details omitted here. Sect. 41.3 is based on appendix E of ref. [21]. We follow Franz [5, 6] in applying the Watson contour method [19] to (41.23). The Airy and Debye approximations to the Hankel functions are given in ref. [1], the Airy expansion of the 1-disk zeros can be found in ref. [7], see also ref. [18] for the expression of $\alpha_\ell(ka)$ to leading order. (For details see refs. [5–7, 20, 21].) That the interior domains of the disks are excluded relatively to the free evolution without scattering obstacles was noted in refs. [3, 16].

The procedure for the semiclassical approximation of a general periodic itinerary (41.20) of period n can be found in ref. [21] for the case of the N -disk systems. The reader interested in the details of the semiclassical reduction is advised to consult this reference.

The ghost orbits were introduced in refs. [2, 3].

Remark 41.2. Krein-Friedel-Lloyd formula. In the literature (see, e.g., refs. [11, 16] based on ref. [2] or ref. [17]) the transition from the quantum mechanics to the semiclassics of scattering problems has been performed via the semiclassical limit of the left hand sides of the Krein-Friedel-Lloyd sum for the (integrated) spectral density [8, 9, 12–15]. See also ref. [4] for a modern discussion of the Krein-Friedel-Lloyd formula and refs. [10, 17] for the connection of (40.17) to the Wigner time delay.

The order of the two limits in (40.18) and (40.17) is essential, see e.g. Balian and Bloch [2] who stress that smoothed level densities should be inserted into the Friedel sums.

The necessity of the $+i\epsilon$ in the semiclassical calculation can be understood by purely phenomenological considerations: Without the $i\epsilon$ term there is no reason why one should be able to neglect spurious periodic orbits which solely are there because of the introduction of the confining boundary. The subtraction of the second (empty) reference system helps just in the removal of those spurious periodic orbits which never encounter the scattering region. The ones that do would still survive the first limit $b \rightarrow \infty$, if they were not damped out by the $+i\epsilon$ term.

exercise 40.1

Remark 41.3. $\mathbf{T}$, $\mathbf{C}^s$, $\mathbf{D}^s$ and $\mathbf{A}^{ss'}$ matrices are trace-class In refs. [21] it has explicitly been shown that the $\mathbf{T}$ -matrix as well as the $\mathbf{C}^s$, $\mathbf{D}^s$ and $\mathbf{A}^{ss'}$ -matrices of the

scattering problem from $N < \infty$ non-overlapping finite disks are all trace-class. The corresponding properties for the single-disk systems is particularly easy to prove.

References

- [1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, 3rd ed. (Dover, New York, 1965).
- [2] R. Balian and C. Bloch, “Solution of the Schrödinger equation in terms of classical paths”, *Ann. Phys.* **85**, 514–545 (1974).
- [3] M. V. Berry, “Quantizing a classically ergodic system: Sinai’s billiard and the KKR method”, *Ann. Phys.* **131**, 163–216 (1981).
- [4] J. S. Faulkner, “Scattering theory and cluster calculations”, *J. Phys. C* **10**, 4661 (1977).
- [5] W. Franz, “Über die Greenschen Funktionen des Zylinders und der Kugel”, *Z. Naturf. A* **9**, 705 (1954).
- [6] W. Franz, *Theorie der Beugung Elektromagnetischer Wellen* (Springer, Berlin, 1957).
- [7] W. Franz and R. Galle, “Semiasymptotische Reihen für die Beugung einer ebenen Welle am Zylinder”, *Z. Naturf. A* **10**, 374 (1955).
- [8] J. Friedel, “XIV. The distribution of electrons round impurities in monovalent metals”, *Phil. Mag.* **43**, 153–189 (1952).
- [9] J. Friedel, “Metallic alloys”, *Nuovo Cim.* **7**, 287–311 (1958).
- [10] P. Gaspard, Scattering resonances: classical and quantum dynamics, in *Quantum Chaos*, Vol. 19, edited by G. Casati, I. Guarneri, and U. Smilansky, Proc. Int. School of Physics Enrico Fermi (1993).
- [11] P. Gaspard and S. A. Rice, “Semiclassical quantization of the scattering from a classically chaotic repeller”, *J. Chem. Phys.* **90**, 2242–2254 (1989).
- [12] M. G. Krein, “On the trace formula in perturbation theory”, *Mat. Sb. (N.S.)* **33**, 597–626 (1953).
- [13] M. G. Krein, “Perturbation determinants and formula for traces of unitary and self-adjoint operators”, *Sov. Math. Dokl.* **3**, 707–710 (1962).
- [14] P. Lloyd, “Wave propagation through an assembly of spheres: II. The density of single-particle eigenstates”, *Proc. Phys. Soc.* **90**, 207–216 (1967).
- [15] P. Lloyd and P. V. Smith, “Multiple scattering theory in condensed materials”, *Adv. Phys.* **21**, 69–142 (1972).
- [16] P. Scherer, Quantenzustände eines Klassisch Chaotischen Billards, PhD thesis (Univ. Köln, 1991).
- [17] W. Thirring, *Quantum Mechanics of Atoms and Molecules*, Vol. 3, Page 48, Sects. 2.3.21 and 2.3.22. (Springer, New York, 1979).
- [18] G. Vattay, A. Wirzba, and P. E. Rosenqvist, “Periodic orbit theory of diffraction”, *Phys. Rev. Lett.* **73**, 2304–2307 (1994).

- [19] G. N. Watson, “The diffraction of electric waves by the earth”, *Proc. R. Soc. Lond. A* **95**, 83–99 (1918).
- [20] A. Wirzba, “Validity of the semiclassical periodic orbit approximation in the two- and three-disk problems”, *Chaos* **2**, 77–83 (1992).
- [21] A. Wirzba, “Quantum mechanics and semiclassics of hyperbolic n-disk scattering systems”, *Phys. Rep.* **309**, 1–116 (1999).

Chapter 42

Helium atom

“But,” Bohr protested, “nobody will believe me unless I can explain every atom and every molecule.” Rutherford was quick to reply, “Bohr, you explain hydrogen and you explain helium and everybody will believe the rest.”

—John Archibald Wheeler (1986)

(G. Tanner)

SO FAR much has been said about 1-dimensional maps, game of pinball and other curious but rather idealized dynamical systems. If you have become impatient and started wondering what good are the methods learned so far in solving real physical problems, we have good news for you. We will show in this chapter that the concepts of symbolic dynamics, unstable periodic orbits, and cycle expansions are essential tools to understand and calculate deterministic and quantum mechanical properties of nothing less than the helium, a dreaded three-body Coulomb problem.

This sounds almost like one step too much at a time; we all know how rich and complicated the dynamics of the three-body problem is – can we really jump from three static disks directly to three charged particles moving under the influence of their mutually attracting or repelling forces? It turns out, we can, but we have to do it with care. The full problem is indeed not accessible in all its detail, but we are able to analyze a somewhat simpler subsystem – collinear helium. This system plays an important role in the deterministic dynamics of the full three-body problem and its quantum spectrum.

The main work in reducing the quantum mechanics of helium to a semiclassical treatment of collinear helium lies in understanding why we are allowed to do so. We will not worry about this too much in the beginning; after all, 80 years and many failed attempts separate Heisenberg, Bohr and others in the 1920ties from the insights we have today on the role chaos plays for helium and its quantum spectrum. We have introduced collinear helium and learned how to integrate its

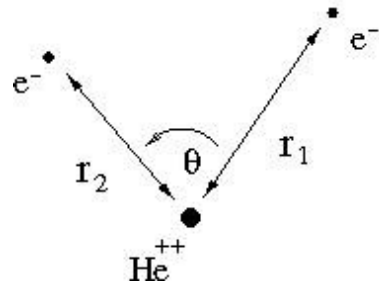


Figure 42.1: Coordinates for the helium three body problem in the plane.

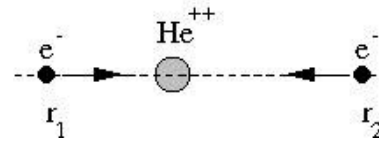


Figure 42.2: Collinear helium, with the two electrons on opposite sides of the nucleus.

trajectories in sect. A2.2. Here we will find periodic orbits and determine the relevant eigenvalues of the Jacobian matrix in sect. 42.1. We will explain in sect. 42.5 why a quantization of the collinear dynamics in helium will enable us to find parts of the full helium spectrum; we then set up the semiclassical spectral determinant and evaluate its cycle expansion. A full quantum justification of this treatment of helium is briefly discussed in sect. 42.5.1.

42.1 Deterministic dynamics of collinear helium

Recapitulating briefly what we learned in sect. A2.2: the collinear helium system consists of two electrons of mass m_e and charge $-e$ moving on a line with respect to a fixed positively charged nucleus of charge $+2e$, as in figure 42.2.

The Hamiltonian can be brought to a non-dimensionalized form

$$H = \frac{p_1^2}{2} + \frac{p_2^2}{2} - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_1 + r_2} = -1. \quad (42.1)$$

The case of negative energies chosen here is the most interesting one for us. It exhibits chaos, unstable periodic orbits and is responsible for the bound states and resonances of the quantum problem treated in sect. 42.5.

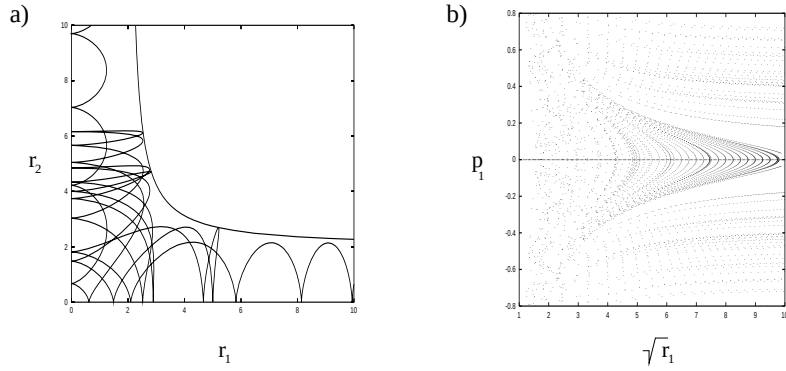
There is another deterministic quantity important for a semiclassical treatment of quantum mechanics, and which will also feature prominently in the discussion in the next section; this is the deterministic action (38.15) which scales with energy as

$$S(E) = \oint d\mathbf{q}(E) \cdot \mathbf{p}(E) = \frac{e^2 m_e^{1/2}}{(-E)^{1/2}} S, \quad (42.2)$$

with S being the action obtained from (42.1) for $E = -1$, and coordinates $\mathbf{q} = (r_1, r_2)$, $\mathbf{p} = (p_1, p_2)$. For the Hamiltonian (42.1), the period of a cycle and its action are related by (38.17),

$$S_p = 2 T_p. \quad (42.3)$$

Figure 42.3: (a) A typical trajectory in the $r_1 - r_2$ plane; the trajectory enters here along the r_1 axis and escapes to infinity along the r_2 axis; (b) return map ($r_2=0$) for collinear helium. Strong chaos prevails for small r_1 near the nucleus.



After a Kustaanheimo–Stiefel transformation

$$r_1 = Q_1^2, \quad r_2 = Q_2^2, \quad p_1 = \frac{P_1}{2Q_1}, \quad p_2 = \frac{P_2}{2Q_2}, \quad (42.4)$$

and reparametrization of time by $d\tau = dt/r_1 r_2$, the equations of motion take form (A2.7)

exercise 42.1

$$\begin{aligned} \dot{P}_1 &= 2Q_1 \left[2 - \frac{P_2^2}{8} - Q_2^2 \left(1 + \frac{Q_2^2}{R_{12}^4} \right) \right]; & \dot{Q}_1 &= \frac{1}{4} P_1 Q_2^2 \\ \dot{P}_2 &= 2Q_2 \left[2 - \frac{P_1^2}{8} - Q_1^2 \left(1 + \frac{Q_1^2}{R_{12}^4} \right) \right]; & \dot{Q}_2 &= \frac{1}{4} P_2 Q_1^2. \end{aligned} \quad (42.5)$$

Individual electron–nucleus collisions at $r_1 = Q_1^2 = 0$ or $r_2 = Q_2^2 = 0$ no longer pose a problem to a numerical integration routine. The equations (A2.7) are singular only at the triple collision $R_{12} = 0$, i.e., when both electrons hit the nucleus at the same time.

The new coordinates and the Hamiltonian (A2.6) are very useful when calculating trajectories for collinear helium; they are, however, less intuitive as a visualization of the three-body dynamics. We will therefore refer to the old coordinates r_1, r_2 when discussing the dynamics and the periodic orbits.

42.2 Chaos, symbolic dynamics and periodic orbits

Let us have a closer look at the dynamics in collinear helium. The electrons are attracted by the nucleus. During an electron–nucleus collision momentum is transferred between the inner and outer electron. The inner electron has a maximal screening effect on the charge of the nucleus, diminishing the attractive force on the outer electron. This electron – electron interaction is negligible if the outer electron is far from the nucleus at a collision and the overall dynamics is regular like in the 1-dimensional Kepler problem.

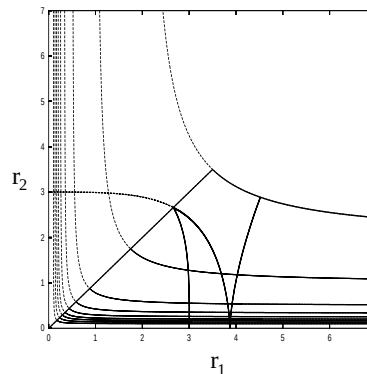


Figure 42.4: The cycle 011 in the fundamental domain $r_1 \geq r_2$ (full line) and in the full domain (dashed line).

Things change drastically if both electrons approach the nucleus nearly simultaneously. The momentum transfer between the electrons depends now sensitively on how the particles approach the origin. Intuitively, these nearly missed triple collisions render the dynamics chaotic. A typical trajectory is plotted in figure 42.3 (a) where we used r_1 and r_2 as the relevant axis. The dynamics can also be visualized in a Poincaré section, see figure 42.3 (b). We plot here the coordinate and momentum of the outer electron whenever the inner particle hits the nucleus, i.e., r_1 or $r_2 = 0$. As the unstructured gray region of the Poincaré section for small r_1 illustrates, the dynamics is chaotic whenever the outer electron is close to the origin during a collision. Conversely, regular motions dominate whenever the outer electron is far from the nucleus. As one of the electrons escapes for almost any starting condition, the system is unbounded: one electron (say electron 1) can escape, with an arbitrary amount of kinetic energy taken by the fugitive. The remaining electron is trapped in a Kepler ellipse with total energy in the range $[-1, -\infty]$. There is no energy barrier which would separate the bound from the unbound regions of the phase space. From general kinematic arguments one deduces that the outer electron will not return when $p_1 > 0$, $r_2 \leq 2$ at $p_2 = 0$, the turning point of the inner electron. Only if the two electrons approach the nucleus almost symmetrically along the line $r_1 = r_2$, and pass close to the triple collision can the momentum transfer between the electrons be large enough to kick one of the particles out completely. In other words, the electron escape originates from the near triple collisions.

The collinear helium dynamics has some important properties which we now list.

42.2.1 Z_2 reflection symmetry

The Hamiltonian (A2.2) is invariant with respect to electron–electron exchange; this Z_2 symmetry corresponds to the mirror symmetry of the potential along the line $r_1 = r_2$, figure 42.4. As a consequence, we can restrict ourselves to the dynamics in the *fundamental domain* $r_1 \geq r_2$ and treat a crossing of the diagonal $r_1 = r_2$ as a hard wall reflection. The dynamics in the full domain can then be reconstructed by unfolding the trajectory through back-reflections. As explained in chapter 25, the dynamics in the fundamental domain is the key to the factorization

of spectral determinants, to be implemented here in (42.16). Note also the similarity between the fundamental domain of the collinear potential figure 42.4, and the fundamental domain figure 15.12 (b) in the 3-disk system, a simpler problem with the same binary symbolic dynamics.



in depth:
sect. 25.6, p. 511

42.2.2 Symbolic dynamics

We have already made the claim that the triple collisions render the collinear helium fully chaotic. We have no proof of the assertion, but the analysis of the symbolic dynamics lends further credence to the claim.

The potential in (42.1) forms a ridge along the line $r_1 = r_2$. One can show that a trajectory passing the ridge must go through at least one two-body collision $r_1 = 0$ or $r_2 = 0$ before coming back to the diagonal $r_1 = r_2$. This suggests a *binary* symbolic dynamics corresponding to the dynamics in the fundamental domain $r_1 \geq r_2$; the symbolic dynamics is linked to the return map $r_2 = 0$ and the symbols 0 and 1 are defined as

- 0: if the trajectory is not reflected from the line $r_1 = r_2$ between two collisions with the nucleus $r_2 = 0$;
- 1: if a trajectory is reflected from the line $r_1 = r_2$ between two collisions with the nucleus $r_2 = 0$.

Empirically, the symbolic dynamics is complete for a return map in the fundamental domain, i.e., there exists a one-to-one correspondence between binary symbol sequences and collinear trajectories in the fundamental domain, with exception of the $\bar{0}$ cycle.

42.2.3 Periodic orbits

The existence of a binary symbolic dynamics makes it easy to count the number of periodic orbits in the fundamental domain, as in sect. 18.7.2. However, mere existence of these cycles does not suffice to calculate semiclassical spectral determinants. We need to determine their phase space trajectories and calculate their periods, topological indices and stabilities. A restriction of the periodic orbit search to a suitable Poincaré section, i.e., $r_2 = 0$ or $r_1 = r_2$, leaves us in general with a 2-dimensional search. Methods to find periodic orbits in multi-dimensional spaces have been described in chapter 16. They depend sensitively on good starting guesses. A systematic search for all orbits can be achieved only after combining multi-dimensional Newton methods with interpolation algorithms

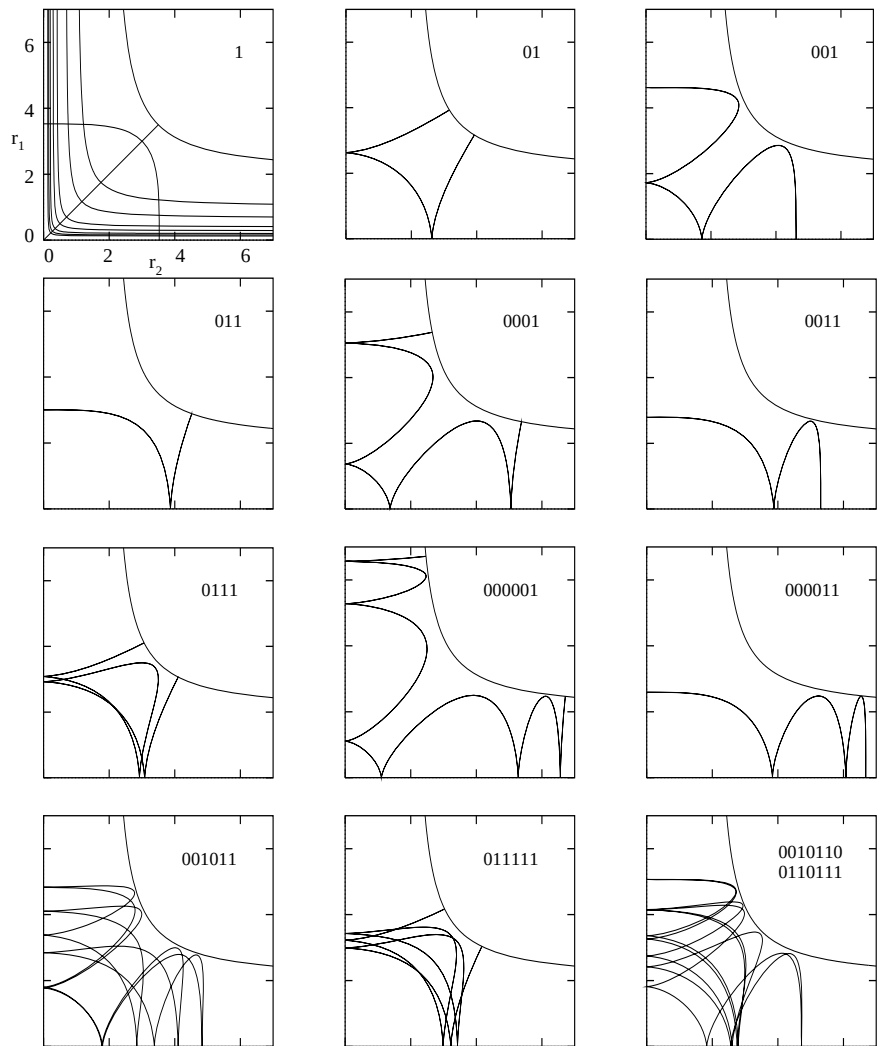


Figure 42.5: Some of the shortest cycles in collinear helium. The deterministic collinear electron motion is bounded by the potential barrier $-1 = -2/r_1 - 2/r_2 + 1/(r_1 + r_2)$ and the condition $r_i \geq 0$. Orbits are shown in the full r_1 - r_2 domain; itineraries refer to dynamics in the $r_1 \geq r_2$ fundamental domain. The last figure, the 14-cycle 00101100110111, is an example of a typical cycle with no symmetry.

based on the binary symbolic dynamics phase space partitioning. All cycles up to symbol length 16 (some 8000 prime cycles) have been computed by such methods, with some examples shown in figure 42.5. All numerical evidence indicates that the dynamics of collinear helium is hyperbolic, and that all periodic orbits are unstable.

Note that the fixed point $\bar{0}$ cycle is not in this list. The $\bar{0}$ cycle would correspond to the situation where the outer electron sits at rest infinitely far from the nucleus while the inner electron bounces back and forth into the nucleus. The orbit is the limiting case of an electron escaping to infinity with zero kinetic energy. The orbit is in the regular (i.e., separable) limit of the dynamics and is thus marginally stable. The existence of this orbit is also related to intermittent behavior generating the quasi-regular dynamics for large r_1 that we have already noted in figure 42.3 (b).

Search algorithm for an arbitrary periodic orbit is quite cumbersome to program. There is, however, a class of periodic orbits, orbits with symmetries, which can be easily found by a one-parameter search. The only symmetry left for the

dynamics in the fundamental domain is time reversal symmetry; a time reversal symmetric periodic orbit is an orbit whose trajectory in phase space is mapped onto itself when changing $(p_1, p_2) \rightarrow (-p_1, -p_2)$, by reversing the direction of the momentum of the orbit. Such an orbit must be a “libration” or self-retracing cycle, an orbit that runs back and forth along the same path in the (r_1, r_2) plane. The cycles $\overline{1}$, $\overline{01}$ and $\overline{001}$ in figure 42.5 are examples of self-retracing cycles. Luckily, the shortest cycles that we desire most ardently have this symmetry.

Why is this observation helpful? A self-retracing cycle must start perpendicular to the boundary of the fundamental domain, that is, on either of the axis $r_2 = 0$ or $r_1 = r_2$, or on the potential boundary $-\frac{2}{r_1} - \frac{2}{r_2} + \frac{1'}{r_1+r_2} = -1$. By shooting off trajectories perpendicular to the boundaries and monitoring the orbits returning to the boundary with the right symbol length we will find time reversal symmetric cycles by varying the starting point on the boundary as the only parameter. But how can we tell whether a given cycle is self-retracing or not? All the relevant information is contained in the itineraries; a cycle is self-retracing if its itinerary is invariant under time reversal symmetry (i.e., read backwards) and a suitable number of cyclic permutations. All binary strings up to length 5 fulfill this condition. The symbolic dynamics contains even more information; we can tell at which boundary the total reflection occurs. One finds that an orbit starts out perpendicular

- to the diagonal $r_1 = r_2$ if the itinerary is time reversal invariant and has an odd number of 1's; an example is the cycle $\overline{001}$ in figure 42.5;
- to the axis $r_2 = 0$ if the itinerary is time reversal invariant and has an even number of symbols; an example is the cycle $\overline{0011}$ in figure 42.5;
- to the potential boundary if the itinerary is time reversal invariant and has an odd number of symbols; an example is the cycle $\overline{011}$ in figure 42.5.

All cycles up to symbol length 5 are time reversal invariant, the first two non-time reversal symmetric cycles are the cycle $\overline{001011}$ in figure 42.5 and, overlying it in this projection, $\overline{001101}$. Their determination requires a two-parameter search. The two cycles are mapped onto each other by time reversal symmetry, i.e., they have the same trace in the r_1 - r_2 plane, but they trace out distinct cycles in the full phase space.

We are ready to integrate trajectories for deterministic collinear helium with the help of the equations of motions (A2.7) and to find all cycles up to period 5. There is only one thing not yet in place; we need the governing equations for the matrix elements of the Jacobian matrix along a trajectory in order to calculate stability indices. We will provide the main equations in the next section, with the details of the derivation relegated to the appendix A8.2.

exercise 42.5

42.3 Local coordinates, Jacobian matrix

In this section, we will derive the equations of motion for the Jacobian matrix along a collinear helium trajectory. The Jacobian matrix is 4-dimensional; the two trivial eigenvectors corresponding to the conservation of energy and displacements along a trajectory can, however, be projected out by suitable orthogonal coordinates transformations, see appendix A4. We will give the transformation to local coordinates explicitly, here for the regularized coordinates (A2.5), and state the resulting equations of motion for the reduced $[2 \times 2]$ Jacobian matrix.

The vector locally parallel to the trajectory is pointing in the direction of the phase space velocity (8.3)

$$v_m = \dot{x}_m(t) = \omega_{mn} \frac{\partial H}{\partial x_n} = (H_{P_1}, H_{P_2}, -H_{Q_1}, -H_{Q_2})^\top,$$

with $H_{Q_i} = \frac{\partial H}{\partial Q_i}$, and $H_{P_i} = \frac{\partial H}{\partial P_i}$, $i = 1, 2$. The vector perpendicular to a trajectory $x(t) = (Q_1(t), Q_2(t), P_1(t), P_2(t))$ and to the energy manifold is given by the gradient of the Hamiltonian (A2.6)

$$\gamma = \nabla H = (H_{Q_1}, H_{Q_2}, H_{P_1}, H_{P_2})^\top.$$

By symmetry $v_m \gamma_m = \omega_{mn} \frac{\partial H}{\partial x_n} \frac{\partial H}{\partial x_m} = 0$, so the two vectors are orthogonal.

Next, we consider the orthogonal matrix

$$\begin{aligned} \mathbf{O} &= (\gamma_1, \gamma_2, \gamma/R, v) \\ &= \begin{pmatrix} -H_{P_2}/R & H_{Q_2} & H_{Q_1}/R & H_{P_1} \\ H_{P_1}/R & -H_{Q_1} & H_{Q_2}/R & H_{P_2} \\ -H_{Q_2}/R & -H_{P_2} & H_{P_1}/R & -H_{Q_1} \\ H_{Q_1}/R & H_{P_1} & H_{P_2}/R & -H_{Q_2} \end{pmatrix} \end{aligned} \quad (42.6)$$

with $R = |\nabla H|^2 = (H_{Q_1}^2 + H_{Q_2}^2 + H_{P_1}^2 + H_{P_2}^2)$, which provides a transformation to local phase space coordinates centered on the trajectory $x(t)$ along the two vectors (γ, v) . The vectors $\gamma_{1,2}$ are phase space vectors perpendicular to the trajectory and to the energy manifold in the 4-dimensional phase space of collinear helium. The Jacobian matrix (4.5) rotated to the local coordinate system by $\mathbf{O}$ then has the form

exercise 42.6

$$\mathbf{m} = \begin{pmatrix} m_{11} & m_{12} & * & 0 \\ m_{21} & m_{22} & * & 0 \\ 0 & 0 & 1 & 0 \\ * & * & * & 1 \end{pmatrix}, \quad M = \mathbf{O}^\top \mathbf{m} \mathbf{O}$$

The linearized motion perpendicular to the trajectory on the energy manifold is described by the $[2 \times 2]$ matrix $\mathbf{m}$; the marginal directions correspond to unit eigenvalues on the diagonal in the 3rd and 4th column and row.

The equations of motion for the reduced Jacobian matrix $\mathbf{m}$ are given by

$$\dot{\mathbf{m}} = \mathbf{1}(t)\mathbf{m}(t), \quad (42.7)$$

with $\mathbf{m}(0) = \mathbf{1}$. The matrix $\mathbf{l}$ depends on the trajectory in phase space and has the form

$$\mathbf{l} = \begin{pmatrix} l_{11} & l_{12} & * & 0 \\ l_{21} & l_{22} & * & 0 \\ 0 & 0 & 0 & 0 \\ * & * & * & 0 \end{pmatrix},$$

where the relevant matrix elements l_{ij} are given by

$$\begin{aligned} l_{11} &= \frac{1}{R} [2H_{Q_1 Q_2} (H_{Q_2} H_{P_1} + H_{Q_1} H_{P_2}) \\ &\quad + (H_{Q_1} H_{P_1} - H_{Q_2} H_{P_2}) (H_{Q_1 Q_1} - H_{Q_2 Q_2} - H_{P_1 P_1} + H_{P_2 P_2})] \\ l_{12} &= -2H_{Q_1 Q_2} (H_{Q_1} H_{Q_2} - H_{P_1} H_{P_2}) \\ &\quad + (H_{Q_1}^2 + H_{P_2}^2) (H_{Q_2 Q_2} + H_{P_1 P_1}) + (H_{Q_2}^2 + H_{P_1}^2) (H_{Q_1 Q_1} + H_{P_2 P_2}) \\ l_{21} &= \frac{1}{R^2} [2(H_{Q_1 P_2} + H_{Q_2 P_1}) (H_{Q_2} H_{P_1} + H_{Q_1} H_{P_2}) \\ &\quad - (H_{P_1}^2 + H_{P_2}^2) (H_{Q_1 Q_1} + H_{Q_2 Q_2}) - (H_{Q_1}^2 + H_{Q_2}^2) (H_{P_1 P_1} + H_{P_2 P_2})] \\ l_{22} &= -l_{11}. \end{aligned} \tag{42.8}$$

Here $H_{Q_i Q_j}$, $H_{P_i P_j}$, $i, j = 1, 2$ are the second partial derivatives of H with respect to the coordinates Q_i , P_i , evaluated at the phase space coordinate of the deterministic trajectory.

42.4 Getting ready

Now everything is in place: the regularized equations of motion can be implemented in a Runge–Kutta or any other integration scheme to calculate trajectories. We have a symbolic dynamics and know how many cycles there are and how to find them (at least up to symbol length 5). We know how to compute the Jacobian matrix whose eigenvalues enter the semiclassical spectral determinant (39.12). By (38.17) the action S_p is proportional to the period of the orbit, $S_p = 2T_p$.

There is, however, still a slight complication. Collinear helium is an invariant 4-dimensional subspace of the full helium phase space. If we restrict the dynamics to angular momentum equal zero, we are left with 6 phase space coordinates. That is not a problem when computing periodic orbits, they are oblivious to the other dimensions. However, the Jacobian matrix does pick up extra contributions. When we calculate the Jacobian matrix for the full problem, we must also allow for displacements out of the collinear plane, so the full Jacobian matrix for dynamics for $L = 0$ angular momentum is 6 dimensional. Fortunately, the linearized dynamics in and off the collinear helium subspace decouple, and the Jacobian matrix can be written in terms of two distinct $[2 \times 2]$ matrices, with trivial eigen-directions providing the remaining two dimensions. The submatrix related to displacements off the linear configuration characterizes the linearized dynamics in the additional degree of freedom, the Θ -coordinate in figure 42.1. It turns out that the linearized

Table 42.1: Action S_p (in units of 2π), stability exponent $\ln |\Lambda_p|/T_p$ for the motion in the collinear plane, winding number σ_p for the motion perpendicular to the collinear plane, and the topological index m_p for all fundamental domain cycles up to topological period 6. Note that by (42.3), $S_p = 2T_p$.

p	$S_p/2\pi$	$\ln \Lambda_p $	σ_p	m_p
1	1.82900	0.6012	0.5393	2
01	3.61825	1.8622	1.0918	4
001	5.32615	3.4287	1.6402	6
011	5.39451	1.8603	1.6117	6
0001	6.96677	4.4378	2.1710	8
0011	7.04134	2.3417	2.1327	8
0111	7.25849	3.1124	2.1705	8
00001	8.56618	5.1100	2.6919	10
00011	8.64306	2.7207	2.6478	10
00101	8.93700	5.1562	2.7291	10
00111	8.94619	4.5932	2.7173	10
01011	9.02689	4.1765	2.7140	10
01111	9.07179	3.3424	2.6989	10
000001	10.13872	5.6047	3.2073	12
000011	10.21673	3.0323	3.1594	12
000101	10.57067	6.1393	3.2591	12
000111	10.57628	5.6766	3.2495	12
001011	10.70698	5.3251	3.2519	12
001101	10.70698	5.3251	3.2519	12
001111	10.74303	4.3317	3.2332	12
010111	10.87855	5.0002	3.2626	12
011111	10.91015	4.2408	3.2467	12

dynamics in the Θ coordinate is stable, corresponding to a bending type motion of the two electrons. We will need the Floquet exponents for all degrees of freedom in evaluating the semiclassical spectral determinant in sect. 42.5.

The numerical values of the actions, Floquet exponents, stability angles, and topological indices for the shortest cycles are listed in table 42.1. These numbers, needed for the semiclassical quantization implemented in the next section, are also helpful in checking your own calculations.

42.5 Semiclassical quantization of collinear helium

Before we get down to a serious calculation of the helium quantum energy levels let us have a brief look at the overall structure of the spectrum. This will give us a preliminary feel for which parts of the helium spectrum are accessible with the help of our collinear model – and which are not. In order to keep the discussion as simple as possible and to concentrate on the semiclassical aspects of our calculations we offer here only a rough overview. For a guide to more detailed accounts see remark 42.4.

42.5.1 Structure of helium spectrum

We start by recalling Bohr's formula for the spectrum of hydrogen like one-electron atoms. The eigenenergies form a Rydberg series

$$E_N = -\frac{e^4 m_e}{\hbar^2} \frac{Z^2}{2N^2}, \quad (42.9)$$

where Ze is the charge of the nucleus and m_e is the mass of the electron. Through the rest of this chapter we adopt the atomic units $e = m_e = \hbar = 1$.

The simplest model for the helium spectrum is obtained by treating the two electrons as independent particles moving in the potential of the nucleus neglecting the electron-electron interaction. Both electrons are then bound in hydrogen like states; the inner electron will see a charge $Z = 2$, screening at the same time the nucleus, the outer electron will move in a Coulomb potential with effective charge $Z - 1 = 1$. In this way obtain a first estimate for the total energy

$$E_{N,n} = -\frac{2}{N^2} - \frac{1}{2n^2} \quad \text{with } n > N. \quad (42.10)$$

This double Rydberg formula contains already most of the information we need to understand the basic structure of the spectrum. The (correct) ionizations thresholds $E_N = -\frac{2}{N^2}$ are obtained in the limit $n \rightarrow \infty$, yielding the ground and excited states of the helium ion He^+ . We will therefore refer to N as the principal quantum number. We also see that all states $E_{N,n}$ with $N \geq 2$ lie above the first ionization threshold for $N = 1$. As soon as we switch on electron-electron interaction these states are no longer bound states; they turn into resonant states which decay into a bound state of the helium ion and a free outer electron. This might not come as a big surprise if we have the deterministic analysis of the previous section in mind: we already found that one of the deterministic electrons will almost always escape after some finite time. More remarkable is the fact that the first, $N = 1$ series consists of true bound states for all n , an effect which can only be understood by quantum arguments.

The hydrogen-like quantum energies (42.9) are highly degenerate; states with different angular momentum but the same principal quantum number N share the same energy. We recall from basic quantum mechanics of hydrogen atom that the possible angular momenta for a given N span $l = 0, 1 \dots N - 1$. How does that affect the helium case? Total angular momentum L for the helium three-body problem is conserved. The collinear helium is a subspace of the deterministic phase space for $L = 0$; we thus expect that we can only quantize helium states corresponding to the total angular momentum zero, a subspectrum of the full helium spectrum. Going back to our crude estimate (42.10) we may now attribute angular momenta to the two independent electrons, l_1 and l_2 say. In order to obtain total angular momentum $L = 0$ we need $l_1 = l_2 = l$ and $l_{z1} = -l_{z2}$, that is, there are N different states corresponding to $L = 0$ for fixed quantum numbers N, n . That means that we expect N different Rydberg series converging to each ionization threshold $E_N = -2/N^2$. This is indeed the case and the N different series can be identified also in the exact helium quantum spectrum, see figure 42.6. The

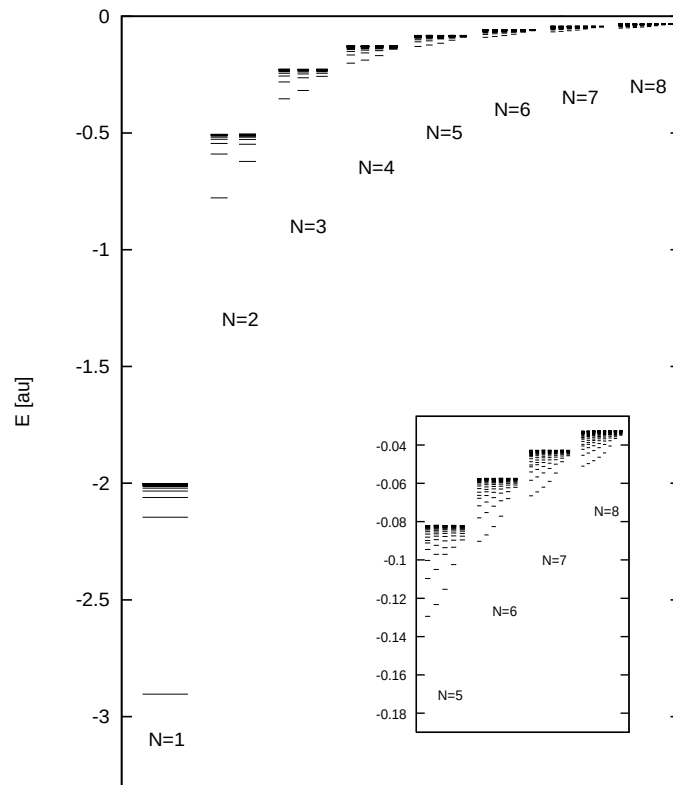


Figure 42.6: The exact quantum helium spectrum for $L = 0$. The energy levels denoted by bars have been obtained from full 3-dimensional quantum calculations [2].

degeneracies between the different N Rydberg series corresponding to the same principal quantum number N , are removed by the electron-electron interaction. We thus already have a rather good idea of the coarse structure of the spectrum.

In the next step, we may even speculate which parts of the $L = 0$ spectrum can be reproduced by the semiclassical quantization of collinear helium. In the collinear helium, both deterministic electrons move back and forth along a common axis through the nucleus, so each has zero angular momentum. We therefore expect that collinear helium describes the Rydberg series with $l = l_1 = l_2 = 0$. These series are the energetically lowest states for fixed (N, n) , corresponding to the Rydberg series on the outermost left side of the spectrum in figure 42.6. We will see in the next section that this is indeed the case and that the collinear model holds down to the $N = 1$ bound state series, including even the ground state of helium! We will also find a semiclassical quantum number corresponding to the angular momentum l and show that the collinear model describes states for mod-

erate angular momentum l as long as $l \ll N$.

remark 42.4

42.5.2 Semiclassical spectral determinant for collinear helium

Nothing but lassitude can stop us now from calculating our first semiclassical eigenvalues. The only thing left to do is to set up the spectral determinant in terms of the periodic orbits of collinear helium and to write out the first few terms of its cycle expansion with the help of the binary symbolic dynamics. The semiclassical spectral determinant (39.12) has been written as product over all cycles of the deterministic systems. The energy dependence in collinear helium enters the classical dynamics only through simple scaling transformations described in sect. A2.2.1 which makes it possible to write the semiclassical spectral determinant in the form

$$\det(\hat{H} - E)_{sc} = \exp \left(- \sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{ir(sS_p - m_p \frac{\pi}{2})}}{(-\det(1 - M_{p\perp}^r))^{1/2} |\det(1 - M_{p\parallel}^r)|^{1/2}} \right), \quad (42.11)$$

with the energy dependence absorbed into the variable

$$s = \frac{e^2}{\hbar} \sqrt{\frac{m_e}{-E}},$$

obtained by using the scaling relation (42.2) for the action. As explained in sect. 42.3, the fact that the $[4 \times 4]$ Jacobian matrix decouples into two $[2 \times 2]$ submatrices corresponding to the dynamics *in* the collinear space and *perpendicular* to it makes it possible to write the denominator in terms of a product of two determinants. Stable and unstable degrees of freedom enter the trace formula in different ways (see (22.9) and the discussion in sect. 23.2.2), reflected by the absence of the modulus sign and the minus sign in front of $\det(1 - M_{\perp})$. The topological index m_p corresponds to the unstable dynamics in the collinear plane. Note that the factor $e^{in\tilde{N}(E)}$ present in (39.12) is absent in (42.11). Collinear helium is an open system, i.e., the eigenenergies are resonances corresponding to the complex zeros of the semiclassical spectral determinant and the mean energy staircase $\tilde{N}(E)$ not defined. In order to obtain a spectral determinant as an infinite product of the form (39.19) we may proceed as in (22.8) by expanding the determinants in (42.11) in terms of the eigenvalues of the corresponding Jacobian matrices. The matrix representing displacements perpendicular to the collinear space has eigenvalues of the form $\exp(\pm 2\pi i \sigma)$, reflecting stable linearized dynamics. σ is the full winding number along the orbit in the stable degree of freedom, multiplicative under multiple repetitions of this orbit. The eigenvalues corresponding to the unstable dynamics along the collinear axis are paired as $\{\Lambda, 1/\Lambda\}$ with $|\Lambda| > 1$ and real. As in (22.8) and (39.19) we may thus write

chapter 40

$$\begin{aligned} & \left[-\det(1 - M_{\perp}^r) |\det(1 - M_{\parallel}^r)| \right]^{-1/2} \\ &= \left[-(1 - \Lambda^r)(1 - \Lambda^{-r})(1 - e^{2\pi i \sigma})(1 - e^{-2\pi i \sigma}) \right]^{-1/2} \\ &= \sum_{k, \ell=0}^{\infty} \frac{1}{|\Lambda^r|^{1/2} \Lambda^{rk}} e^{-ir(\ell+1/2)\sigma}. \end{aligned} \quad (42.12)$$

The $\pm$ sign corresponds to the hyperbolic/inverse hyperbolic periodic orbits with positive/negative eigenvalues Λ . Using the relation (42.13) we see that the sum over r in (42.11) is the expansion of the logarithm, so the semiclassical spectral determinant can be rewritten as a product over dynamical zeta functions, as in (22.8):

$$\det(\hat{H} - E)_{sc} = \prod_{k=0}^{\infty} \prod_{m=0}^{\infty} \zeta_{k,m}^{-1} = \prod_{k=0}^{\infty} \prod_{m=0}^{\infty} \prod_p (1 - t_p^{(k,m)}), \quad (42.13)$$

where the cycle weights are given by

$$t_p^{(k,m_p)} = \frac{1}{|\Lambda|^{1/2} \Lambda^k} e^{i(sS_p - m_p \frac{\pi}{2} - 4\pi(\ell+1/2)\sigma_p)}, \quad (42.14)$$

and m_p is the topological index for the motion in the collinear plane which equals twice the topological period of the cycle. The two independent directions perpendicular to the collinear axis lead to a twofold degeneracy in this degree of freedom which accounts for an additional factor 2 in front of the winding number σ . The values for the actions, winding numbers and stability indices of the shortest cycles in collinear helium are listed in table 42.1.

The integer indices ℓ and k play very different roles in the semiclassical spectral determinant (42.13). A linearized approximation of the flow along a cycle corresponds to a harmonic approximation of the potential in the vicinity of the trajectory. Stable motion corresponds to a harmonic oscillator potential, unstable motion to an inverted harmonic oscillator. The index ℓ which contributes as a phase to the cycle weights in the dynamical zeta functions can therefore be interpreted as a harmonic oscillator quantum number; it corresponds to vibrational modes in the Θ coordinate and can in our simplified picture developed in sect. 42.5.1 be related to the quantum number $l = l_1 = l_2$ representing the single particle angular momenta. Every distinct ℓ value corresponds to a full spectrum which we obtain from the zeros of the semiclassical spectral determinant $1/\zeta_\ell$ keeping ℓ fixed. The harmonic oscillator approximation will eventually break down with increasing off-line excitations and thus increasing ℓ . The index k corresponds to ‘excitations’ along the unstable direction and can be identified with local resonances of the inverted harmonic oscillator centered on the given orbit. The cycle contributions $t_p^{(k,m)}$ decrease exponentially with increasing k . Higher k terms in an expansion of the determinant give corrections which become important only for large negative imaginary s values. As we are interested only in the leading zeros of (42.13), i.e., the zeros closest to the real energy axis, it is sufficient to take only the $k = 0$ terms into account.

appendix A39

Next, let us have a look at the discrete symmetries discussed in sect. 42.2. Collinear helium has a Z_2 symmetry as it is invariant under reflection across the $r_1 = r_2$ line corresponding to the electron-electron exchange symmetry. As explained in example 25.9 and sect. 25.5, we may use this symmetry to factorize the semiclassical spectral determinant. The spectrum corresponding to the states symmetric or antisymmetric with respect to reflection can be obtained by writing the dynamical zeta functions in the symmetry factorized form

$$1/\zeta^{(\ell)} = \prod_a (1 - t_a)^2 \prod_{\tilde{s}} (1 - t_{\tilde{s}}^2). \quad (42.15)$$

Here, the first product is taken over all asymmetric prime cycles, i.e., cycles that are not self-dual under the Z_2 symmetry. Such cycles come in pairs, as two equivalent orbits are mapped into each other by the symmetry transformation. The second product runs over all self-dual cycles; these orbits cross the axis $r_1 = r_2$ twice at a right angle. The self-dual cycles close in the fundamental domain $r_1 \leq r_2$ already at half the period compared to the orbit in the full domain, and the cycle weights $t_{\bar{s}}$ in (42.15) are the weights of fundamental domain cycles. The Z_2 symmetry now leads to the factorization of (42.15) $1/\zeta = \zeta_+^{-1} \zeta_-^{-1}$, with

section 25.5

$$\begin{aligned} 1/\zeta_+^{(\ell)} &= \prod_a (1 - t_a) \prod_{\bar{s}} (1 - t_{\bar{s}}), \\ 1/\zeta_-^{(\ell)} &= \prod_a (1 - t_a) \prod_{\bar{s}} (1 + t_{\bar{s}}), \end{aligned} \quad (42.16)$$

setting $k = 0$ in what follows. The symmetric subspace resonances are given by the zeros of $1/\zeta_+^{(\ell)}$, antisymmetric resonances by the zeros of $1/\zeta_-^{(\ell)}$, with the two dynamical zeta functions defined as products over orbits in the fundamental domain. The symmetry properties of an orbit can be read off directly from its symbol sequence, as explained in sect. 42.2. An orbit with an odd number of 1's in the itinerary is self-dual under the Z_2 symmetry and enters the spectral determinant in (42.16) with a negative or a positive sign, depending on the symmetry subspace under consideration.

42.5.3 Cycle expansion results

So far we have established a factorized form of the semiclassical spectral determinant and have thereby picked up two *good quantum numbers*; the quantum number m has been identified with an excitation of the bending vibrations, the exchange symmetry quantum number ± 1 corresponds to states being symmetric or antisymmetric with respect to the Z_2 electron-electron exchange. We may now start writing down the binary cycle expansion (23.8) and determine the zeros of spectral determinant. There is, however, still another problem: there is no cycle 0 in the collinear helium. The symbol sequence $\bar{0}$ corresponds to the limit of an outer electron fixed with zero kinetic energy at $r_1 = \infty$, the inner electron bouncing back and forth into the singularity at the origin. This introduces intermittency in our system, a problem discussed in chapter 29. We note that the behavior of cycles going far out in the channel r_1 or $r_2 \rightarrow \infty$ is very different from those staying in the near core region. A cycle expansion using the binary alphabet reproduces states where both electrons are localized in the near core regions: these are the lowest states in each Rydberg series. The states converging to the various ionization thresholds $E_N = -2/N^2$ correspond to eigenfunctions where the wave function of the outer electron is stretched far out into the ionization channel $r_1, r_2 \rightarrow \infty$. To include those states, we have to deal with the dynamics in the limit of large r_1, r_2 . This turns out to be equivalent to switching to a symbolic dynamics with an infinite alphabet of table 29.1, a step beyond our ambition horizon right now. With this observation in mind, we may write the cycle expansion (42.17) for a

section 29.3

remark 42.5

binary alphabet without the $\bar{0}$ cycle as

$$\begin{aligned} 1/\zeta^\ell(s) = & 1 - t_1^{(\ell)} - t_{01}^{(\ell)} - [t_{001}^{(\ell)} + t_{011}^{(\ell)} - t_{01}^{(\ell)} t_1^{(\ell)}] \\ & - [t_{0001}^{(\ell)} + t_{0011}^{(\ell)} - t_{001}^{(\ell)} t_1^{(\ell)} + t_{0111}^{(\ell)} - t_{011}^{(\ell)} t_1^{(\ell)}] - \dots \end{aligned} \quad (42.17)$$

The weights $t_p^{(\ell)}$ are given in (42.13), with contributions of orbits and composite orbits of the same total symbol length collected within square brackets. The cycle expansion depends only on the deterministic actions, stability indices and winding numbers, given for orbits up to period 6 in table 42.1. To get reacquainted with the cycle expansion formula (42.17), consider a truncation of the series after the first term

$$1/\zeta^{(\ell)}(s) \approx 1 - t_1.$$

The quantization condition $1/\zeta^{(\ell)}(s) = 0$ leads to

$$E_{m,N} = -\frac{(S_1/2\pi)^2}{[m + \frac{1}{2} + 2(N + \frac{1}{2})\sigma_1]^2}, \quad m, N = 0, 1, 2, \dots, \quad (42.18)$$

with $S_1/2\pi = 1.8290$ for the action and $\sigma_1 = 0.5393$ for the winding number, see table 42.1, the 1 cycle in the fundamental domain. This cycle can be described as the *asymmetric stretch* orbit, see figure 42.5. The additional quantum number N in (42.18) corresponds to the principal quantum number defined in sect. 42.5.1. The states described by the quantization condition (42.18) are those centered closest to the nucleus and correspond therefore to the lowest states in each Rydberg series (for a fixed m and N values), in figure 42.6. The simple formula (42.18) gives already a rather good estimate for the ground state of helium! Results obtained from (42.18) are tabulated in table 42.2, see the 3rd column under $j = 1$ and the comparison with the full quantum calculations.

In order to obtain higher excited quantum states, we need to include more orbits in the cycle expansion (42.17), covering more of the phase space dynamics further away from the center. Taking longer and longer cycles into account, we indeed reveal more and more states in each N -series for fixed m . This is illustrated by the data listed in table 42.2 for symmetric states obtained from truncations of the cycle expansion of $1/\zeta_+$.

exercise 42.7

Results of the same quality are obtained for antisymmetric states by calculating the zeros of $1/\zeta_-^{(\ell)}$. Repeating the calculation with $\ell = 1$ or higher in (42.16) reveals states in the Rydberg series which are to the right of the energetically lowest series in figure 42.6.

Résumé

We have covered a lot of ground, starting with considerations of the deterministic properties of a three-body Coulomb problem, and ending with the semiclassical

Table 42.2: Collinear helium, real part of the symmetric subspace resonances obtained by a cycle expansion (42.17) up to cycle period j . The exact quantum energies [2] are in the last column. The states are labeled by their principal quantum numbers. A dash as an entry indicates a missing zero at that level of approximation.

N	m	$j = 1$	$j = 4$	$j = 8$	$j = 12$	$j = 16$	$-E_{\text{qm}}$
1	1	3.0970	2.9692	2.9001	2.9390	2.9248	2.9037
2	2	0.8044	0.7714	0.7744	0.7730	0.7727	0.7779
2	3	—	0.5698	0.5906	0.5916	0.5902	0.5899
2	4	—	—	—	0.5383	0.5429	0.5449
3	3	0.3622	0.3472	0.3543	0.3535	0.3503	0.3535
3	4	—	—	0.2812	0.2808	0.2808	0.2811
3	5	—	—	0.2550	0.2561	0.2559	0.2560
3	6	—	—	—	0.2416	0.2433	0.2438
4	4	0.2050	0.1962	0.1980	0.2004	0.2012	0.2010
4	5	—	0.1655	0.1650	0.1654	0.1657	0.1657
4	6	—	—	0.1508	0.1505	0.1507	0.1508
4	7	—	—	0.1413	0.1426	0.1426	0.1426

helium spectrum. We saw that the three-body problem restricted to the dynamics on a collinear appears to be fully chaotic; this implies that traditional semiclassical methods such as WKB quantization will not work and that we need the full periodic orbit theory to obtain leads to the semiclassical spectrum of helium. As a piece of unexpected luck the symbolic dynamics is simple, and the semiclassical quantization of the collinear dynamics yields an important part of the helium spectrum, including the ground state, to a reasonable accuracy. A sceptic might say: “Why bother with all the semiclassical considerations? A straightforward numerical quantum calculation achieves the same goal with better precision.” While this is true, the semiclassical analysis offers new insights into the *structure* of the spectrum. We discovered that the dynamics perpendicular to the collinear plane was stable, giving rise to an additional (approximate) quantum number ℓ . We thus understood the origin of the different Rydberg series depicted in figure 42.6, a fact which is not at all obvious from a numerical solution of the quantum problem.

chapter 37

Having traversed the long road from the deterministic game of pinball all the way to a credible helium spectrum computation, we could declare victory and fold down this enterprise. Nevertheless, there is still much to think about - what about such quintessentially quantum effects as diffraction, tunnelling, ...? As we shall see, the periodic orbit theory has still much of interest to offer.

chapter 43

Commentary

Remark 42.1. Sources. The full 3-dimensional Hamiltonian after elimination of the center of mass coordinates, and an account of the finite nucleus mass effects is given in ref. [8]. The general two-body collision regularizing Kustaanheimo–Stiefel transformation [6], a generalization of Levi-Civita’s [7] Pauli matrix two-body collision regular-

ization for motion in a plane, is due to Kustaanheimo [5] who realized that the correct higher-dimensional generalization of the “square root removal” trick (A2.18), by introducing a vector Q with property $r = |Q|^2$, is the same as Dirac’s trick of getting linear equation for spin 1/2 fermions by means of spinors. Vector spaces equipped with a product and a known satisfy $|Q \cdot Q| = |Q|^2$ define *normed algebras*. They appear in various physical applications - as quaternions, octonions, spinors. The technique was originally developed in celestial mechanics [10] to obtain numerically stable solutions for planetary motions. The basic idea was in place as early as 1931, when H. Hopf [4] used a KS transformation in order to illustrate a Hopf’s invariant. The KS transformation for the collinear helium was introduced in ref. [8].

Remark 42.2. Complete binary symbolic dynamics. No stable periodic orbit and no exception to the binary symbolic dynamics of the collinear helium cycles have been found in numerical investigations. A proof that all cycles are unstable, that they are uniquely labeled by the binary symbolic dynamics, and that this dynamics is complete is, however, still missing. The conjectured Markov partition of the phase space is given by the triple collision manifold, i.e., by those trajectories which start in or end at the singular point $r_1 = r_2 = 0$. See also ref. [8].

Remark 42.3. Spin and particle exchange symmetry. In our presentation of collinear helium we have completely ignored all dynamical effects due to the spin of the particles involved, such as the electronic spin-orbit coupling. Electrons are fermions and that determines the symmetry properties of the quantum states. The total wave function, including the spin degrees of freedom, must be antisymmetric under the electron-electron exchange transformation. That means that a quantum state symmetric in the position variables must have an antisymmetric spin wave function, i.e., the spins are antiparallel and the total spin is zero (singletstate). Antisymmetric states have symmetric spin wave function with total spin 1 (tripletstates). The threefold degeneracy of spin 1 states is lifted by the spin-orbit coupling.

Remark 42.4. Helium quantum numbers. The classification of the helium states in terms of single electron quantum numbers, sketched in sect. 42.5.1, prevailed until the 1960’s; a growing discrepancy between experimental results and theoretical predictions made it necessary to refine this picture. In particular, the different Rydberg series sharing a given N -quantum number correspond, roughly speaking, to a quantization of the inter electronic angle Θ , see figure 42.1, and can not be described in terms of single electron quantum numbers l_1, l_2 . The fact that something is slightly wrong with the single electron picture laid out in sect. 42.5.1 is highlighted when considering the collinear configuration where both electrons are on the *same* side of the nucleus. As both electrons again have angular momentum equal to zero, the corresponding quantum states should also belong to single electron quantum numbers $(l_1, l_2) = (0, 0)$. However, the single electron picture breaks down completely in the limit $\Theta = 0$ where electron-electron interaction becomes the dominant effect. The quantum states corresponding to this deterministic configuration are distinctively different from those obtained from the collinear dynamics with electrons on different sides of the nucleus. The Rydberg series related to the deterministic $\Theta = 0$ dynamics are on the outermost right side in each N subspectrum in figure 42.6, and contain the energetically highest states for given N, n quantum numbers, see also remark 42.5. A detailed account of the historical development as well as a modern interpretation of the spectrum can be found in ref. [11].

Remark 42.5. Beyond the unstable collinear helium subspace. The semiclassical quantization of the chaotic collinear helium subspace is discussed in refs. [1, 3, 14]. Deterministic and semiclassical considerations beyond what has been discussed in sect. 42.5 follow several other directions, all outside the main thrust of ChaosBook.

A deterministic study of the dynamics of collinear helium where both electrons are on the same side of the nucleus reveals that this configuration is fully stable both in the collinear plane and perpendicular to it. The corresponding quantum states can be obtained with the help of an approximate EBK-quantization which reveals helium resonances with extremely long lifetimes (quasi - bound states in the continuum). These states form the energetically highest Rydberg series for a given principal quantum number N , see figure 42.6. Details can be found in refs. [9, 13].

In order to obtain the Rydberg series structure of the spectrum, i.e., the succession of states converging to various ionization thresholds, we need to take into account the dynamics of orbits which make large excursions along the r_1 or r_2 axis. In the chaotic collinear subspace these orbits are characterized by itineraries of form $(a0^n)$ where a stands for an arbitrary binary symbol sequence and 0^n is a succession of n 0's in a row. A summation of the form $\sum_{n=0}^{\infty} t_a 0^n$, where t_p are the cycle weights in (42.13), and cycle expansion of indeed yield all Rydberg states up the various ionization thresholds, see ref. [12]. For a comprehensive overview on spectra of two-electron atoms and semiclassical treatments ref. [11].

References

- [1] R. Blümel and W. P. Reinhardt, “Where is the chaos in two-electron atoms?”, in *Directions in Chaos*, Vol. 4, edited by D. H. Feng and J.-M. Yuan (World Scientific, Singapore, 1992), pp. 245–319.
- [2] A. Burgers, D. Wintgen, and J.-M. Rost, “Highly doubly excited S states of the helium atom”, *J. Phys. B* **28**, 3163 (1995).
- [3] G. S. Ezra, K. Richter, G. Tanner, and D. Wintgen, “Semiclassical cycle expansion for the helium atom”, *J. Phys. B* **24**, L413–L420 (1991).
- [4] H. Hopf, “Über die Abbildungen der dreidimensionalen Sphäre auf die Kugelfläche”, *Math. Ann.* **104**, 637–665 (1931).
- [5] P. Kustaanheimo, “Spinor regularization of the Kepler motion”, *Ann. Univ. Turku, Ser. AI.* **73**, 204–219 (1964).
- [6] P. Kustaanheimo and E. Stiefel, “Perturbation theory of Kepler motion based on spinor regularization”, *J. Ang. Math* **218**, 204–219 (1965).
- [7] T. Levi-Civita, “Sur la résolution qualitative du problème restreint des trois corps”, *Acta Math.* **30**, 305–327 (1906).
- [8] K. Richter, G. Tanner, and D. Wintgen, “Classical mechanics of two-electron atoms”, *Phys. Rev. A* **48**, 4182–4196 (1993).
- [9] K. Richter and D. Wintgen, “Calculations of planetary atom states”, *J. Phys. B* **24**, L565 (1991).

- [10] E. L. Stiefel and G. Scheifele, *Linear and Regular Celestial Mechanics* (Springer, New York, 1971).
- [11] G. Tanner, K. Richter, and J.-M. Rost, “The theory of two-electron atoms: between ground state and complete fragmentation”, *Rev. Mod. Phys.* **72**, 497–544 (2000).
- [12] G. Tanner and D. Wintgen, “Semiclassical quantization of intermittency in helium”, *Phys. Rev. Lett.* **75**, 2928–2931 (1995).
- [13] D. Wintgen and K. Richter, “Semiclassical nature of planetary atom states”, *Comments At. Mol. Phys.* **29**, 261–273 (1994).
- [14] D. Wintgen, K. Richter, and G. Tanner, “The semiclassical helium atom”, *Chaos* **2**, 19–33 (1992).

42.6 Exercises

- 42.1. **Kustaanheimo–Stiefel transformation.** Check the Kustaanheimo–Stiefel regularization for collinear helium; derive the Hamiltonian (A2.6) and the collinear helium equations of motion (A2.7).
- 42.2. **Helium in the plane.** Starting with the helium Hamiltonian in the infinite nucleus mass approximation $m_{he} = \infty$, and angular momentum $L = 0$, show that the three body problem can be written in terms of three independent coordinates only, the electron-nucleus distances r_1 and r_2 and the inter-electron angle Θ , see figure A2.1.
- 42.3. **Helium trajectories.** Do some trial integrations of the collinear helium equations of motion (A2.7). Due to the energy conservation, only three of the phase space coordinates (Q_1, Q_2, P_1, P_2) are independent. Alternatively, you can integrate in 4 dimensions and use the energy conservation as a check on the quality of your integrator.
- The dynamics can be visualized as a motion in the original configuration space (r_1, r_2) , $r_i \geq 0$ quadrant, or, better still, by an appropriately chosen 2-dimensional Poincaré section, exercise 42.4. Most trajectories will run away, do not be surprised - the deterministic collinear helium is unbound. Try to guess approximately the shortest cycle of figure 42.4.
- 42.4. **A Poincaré section for collinear Helium.** Construct a Poincaré section of figure 42.3b that reduces the helium flow to a map. Try to delineate regions which correspond to finite symbol sequences, i.e. initial conditions that follow the same topological itinerary in figure 42.3a space for a finite number of bounces. Such rough partition can be used to initiate 2-dimensional Newton-Raphson method searches for helium cycles, exercise 42.5.
- 42.5. **Collinear helium cycles.** The motion in the (r_1, r_2) plane is topologically similar to the pinball motion in a 3-disk system, except that the motion is in the Coulomb potential.
- Just as in the 3-disk system the dynamics is simplified if viewed in the *fundamental domain*, in this case the region between r_1 axis and the $r_1 = r_2$ diagonal. Modify your integration routine so the trajectory bounces off the diagonal as off a mirror. Miraculously, the symbolic dynamics for the survivors again turns out to be binary, with 0 symbol signifying a bounce off the r_1 axis, and 1 symbol for a bounce off the diagonal. Just as in the 3-disk game of pinball, we thus know what cycles need to be computed for the cycle expansion (42.17).
- Guess some short cycles by requiring that topologically they correspond to sequences of bounces either returning to the same r_i axis or reflecting off the diagonal. Now either Use special symmetries of orbits such as self-retracing to find all orbits up to period 5 by a 1-dimensional Newton search.
- 42.6. **Collinear helium cycle stabilities.** Compute the eigenvalues for the cycles you found in exercise 42.5, as described in sect. 42.3. You may either integrate the reduced 2×2 matrix using equations (42.7) together with the generating function $\mathbf{I}$ given in local coordinates by (42.8) or integrate the full 4×4 Jacobian matrix, see appendix A8.2. Integration in 4 dimensions should give eigenvalues of the form $(1, 1, \Lambda_p, 1/\Lambda_p)$; The unit eigenvalues are due to the usual periodic orbit invariances; displacements along the orbit as well as perpendicular to the energy manifold are conserved; the latter one provides a check of the accuracy of your computation. Compare with table 42.1; you should get the actions and Lyapunov exponents right, but topological indices and stability angles we take on faith.
- 42.7. **Helium eigenenergies.** Compute the lowest eigenenergies of singlet and triplet states of helium by substituting cycle data into the cycle expansion (42.17) for the symmetric and antisymmetric zeta functions (42.16). Probably the quickest way is to plot the magnitude of

the zeta function as function of real energy and look for the minima. As the eigenenergies in general have a small imaginary part, a contour plot such as figure 23.1, can yield informed guesses. Better way would be to find the zeros by Newton method, sect. 23.1.

Chapter 43

Diffraction distraction

(N. Whelan)

DIFFRACTION EFFECTS characteristic to scattering off wedges are incorporated into the periodic orbit theory.

43.1 Quantum eavesdropping

As noted in chapter 42, the deterministic mechanics of the helium atom is undefined at the instant of a triple collision. This is a common phenomenon - there is often some singularity or discontinuity in the deterministic mechanics of physical systems. This discontinuity can even be helpful in classifying the dynamics. The points in phase space which have a past or future at the discontinuity form manifolds which divide the phase space and provide the symbolic dynamics. The general rule is that quantum mechanics smoothes over these discontinuities in a process we interpret as diffraction. We solve the local diffraction problem quantum mechanically and then incorporate this into our global solution. By doing so, we reconfirm the central leitmotif of this treatise: think locally - act globally.

While being a well-motivated physical example, the helium atom is somewhat involved. In fact, so involved that we do not have a clue how to do it. In its place we illustrate the concept of diffractive effects with a pinball game. There are various classes of discontinuities which a billiard can have. There may be a grazing condition such that some trajectories hit a smooth surface while others are unaffected - this leads to the creeping described in chapter 40. There may be a vertex such that trajectories to one side bounce differently from those to the other side. There may be a point scatterer or a magnetic flux line such that we do not know how to continue deterministic mechanics through the discontinuities. In what follows, we specialize the discussion to the second example - that of vertices or wedges. To further simplify the discussion, we consider the special case of a half line which can be thought of as a wedge of angle zero.

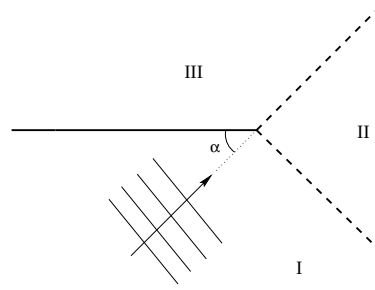


Figure 43.1: Scattering of a plane wave off a half line.

We start by solving the problem of the scattering of a plane wave off a half line (see figure 43.1). This is the local problem whose solution we will use to construct a global solution of more complicated geometries. We define the vertex to be the origin and launch a plane wave at it from an angle α . What is the total field? This is a problem solved by Sommerfeld [17–19] in 1896 and our discussion closely follows his.

The total field consists of three parts - the incident field, the reflected field and the diffractive field. Ignoring the third of these for the moment, we see that the space is divided into three regions. In region I there is both an incident and a reflected wave. In region II there is only an incident field. In region III there is nothing so we call this the shadowed region. However, because of diffraction the field does enter this region. This accounts for why you can overhear a conversation if you are on the opposite side of a thick wall but with a door a few meters away. Traditionally such effects have been ignored in semiclassical calculations because they are relatively weak. However, they can be significant.

To solve this problem Sommerfeld worked by analogy with the full line case, so let us briefly consider that much simpler problem. There we know that the problem can be solved by images. An incident wave of amplitude A is of the form

$$v(r, \psi) = Ae^{-ikr \cos \psi} \quad (43.1)$$

where $\psi = \phi - \alpha$ and ϕ is the angular coordinate. The total field is then given by the method of images as

$$v_{\text{tot}} = v(r, \phi - \alpha) - v(r, \phi + \alpha), \quad (43.2)$$

where the negative sign ensures that the boundary condition of zero field on the line is satisfied.

Sommerfeld then argued that $v(r, \psi)$ can also be given a complex integral representation

$$v(r, \psi) = A \int_C d\beta f(\beta, \psi) e^{-ikr \cos \beta}. \quad (43.3)$$

This is certainly correct if the function $f(\beta, \psi)$ has a pole of residue $1/2\pi i$ at $\beta = -\psi$ and if the contour C encloses that pole. One choice is

$$f(\beta, \psi) = \frac{1}{2\pi} \frac{e^{i\beta}}{e^{i\beta} - e^{-i\psi}}. \quad (43.4)$$

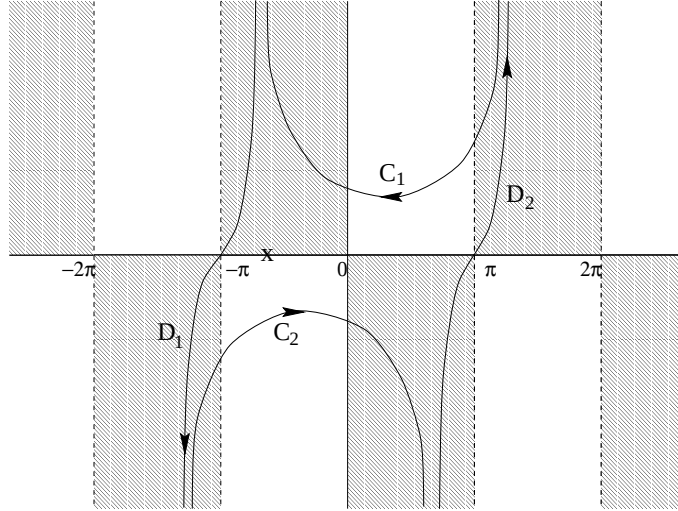


Figure 43.2: The contour in the complex β plane. The pole is at $\beta = -\psi$ (marked by $\times$ in the figure) and the integrand approaches zero in the shaded regions as the magnitude of the imaginary part of β approaches infinity.

(We choose the pole to be at $\beta = -\psi$ rather than $\beta = \psi$ for reasons discussed later.) One valid choice for the contour is shown in figure 43.2. This encloses the pole and vanishes as $|\text{Im}\beta| \rightarrow \infty$ (as denoted by the shading). The sections D_1 and D_2 are congruent because they are displaced by 2π . However, they are traversed in an opposite sense and cancel, so our contour consists of just the sections C_1 and C_2 . The motivation for expressing the solution in this complicated manner should become clear soon.

What have we done? We extended the space under consideration by a factor of two and then constructed a solution by assuming that there is also a source in the unphysical space. We superimpose the solutions from the two sources and at the end only consider the solution in the physical space to be meaningful. Furthermore, we expressed the solution as a contour integral which reflects the 2π periodicity of the problem. The half line scattering problem follows by analogy.

Whereas for the full line the field is periodic in 2π , for the half line it is periodic in 4π . This can be seen by the fact that the field can be expanded in a series of the form $\{\sin(\phi/2), \sin(\phi), \sin(3\phi/2), \dots\}$. As above, we extend the space by thinking of it as two sheeted. The physical sheet is as shown in figure 43.1 and the unphysical sheet is congruent to it. The sheets are glued together along the half line so that a curve in the physical space which intersects the half line is continued in the unphysical space and vice-versa. The boundary conditions are that the total field is zero on both faces of the half line (which are physically distinct boundary conditions) and that as $r \rightarrow \infty$ the field is composed solely of plane waves and outgoing circular waves of the form $g(\phi) \exp(ikr) / \sqrt{kr}$. This last condition is a result of Huygens' principle.

We assume that the complete solution is also given by the method of images as

$$v_{\text{tot}} = u(r, \phi - \alpha) - u(r, \phi + \alpha). \quad (43.5)$$

where $u(r, \psi)$ is a 4π -periodic function to be determined. The second term is

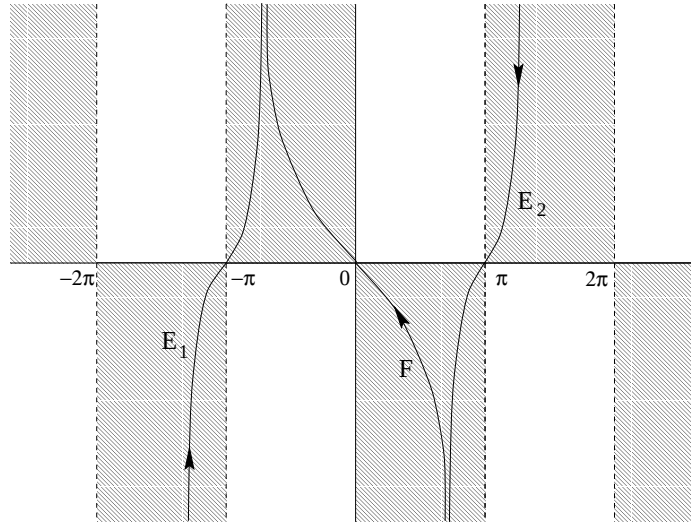


Figure 43.3: The contour used to evaluate the diffractive field after the contribution of possible poles has been explicitly evaluated. The curve F is traversed twice in opposite directions and has no net contribution.

interpreted as an incident field from the unphysical space and the negative sign guarantees that the solution vanishes on both faces of the half line. Sommerfeld then made the ansatz that u is as given in equation (43.3) with the same contour $C_1 + C_2$ but with the 4π periodicity accounted for by replacing equation (43.4) with

$$f(\beta, \psi) = \frac{1}{4\pi} \frac{e^{i\beta/2}}{e^{i\beta/2} - e^{-i\psi/2}}. \quad (43.6)$$

(We divide by 4π rather than 2π so that the residue is properly normalized.) The integral (43.3) can be thought of as a linear superposition of an infinity of plane waves each of which satisfies the Helmholtz equation $(\nabla^2 + k^2)v = 0$, and so their combination also satisfies the Helmholtz equation. We will see that the diffracted field is an outgoing circular wave; this being a result of choosing the pole at $\beta = -\psi$ rather than $\beta = \psi$ in equation (43.4). Therefore, this ansatz is a solution of the equation and satisfies all boundary conditions and therefore constitutes a valid solution. By uniqueness this is the only solution.

In order to further understand this solution, it is useful to massage the contour. Depending on ϕ there may or may not be a pole between $\beta = -\pi$ and $\beta = \pi$. In region I, both functions $u(r, \phi \pm \alpha)$ have poles which correspond to the incident and reflected waves. In region II, only $u(r, \phi - \alpha)$ has a pole corresponding to the incident wave. In region III there are no poles because of the shadow. Once we have accounted for the geometrical waves (i.e., the poles), we extract the diffracted waves by saddle point analysis at $\beta = \pm\pi$. We do this by deforming the contours C so that they go through the saddles as shown in figure 43.2.

Contour C_1 becomes $E_2 + F$ while contour C_2 becomes $E_1 - F$ where the minus sign indicates that it is traversed in a negative sense. As a result, F has no net contribution and the contour consists of just E_1 and E_2 .

As a result of these machinations, the curves E are simply the curves D of figure 43.2 but with a reversed sense. Since the integrand is no longer 2π periodic, the contributions from these curves no longer cancel. We evaluate both stationary

phase integrals to obtain

$$u(r, \psi) \approx -A \frac{e^{i\pi/4}}{\sqrt{8\pi}} \sec(\psi/2) \frac{e^{ikr}}{\sqrt{kr}} \quad (43.7)$$

so that the total diffracted field is

$$v_{\text{diff}} = -A \frac{e^{i\pi/4}}{\sqrt{8\pi}} \left(\sec\left(\frac{\phi - \alpha}{2}\right) - \sec\left(\frac{\phi + \alpha}{2}\right) \right) \frac{e^{ikr}}{\sqrt{kr}}. \quad (43.8)$$

Note that this expression breaks down when $\phi \pm \alpha = \pi$. These angles correspond to the borders among the three regions of figure 43.1 and must be handled more carefully - we can not do a stationary phase integral in the vicinity of a pole. However, the integral representation (43.3) and (43.6) is uniformly valid.

exercise 43.1

We now turn to the simple task of translating this result into the language of semiclassical Green's functions. Instead of an incident plane wave, we assume a source at point x' and then compute the resulting field at the receiver position x . If x is in region I, there is both a direct term, and a reflected term, if x is in region II there is only a direct term and if x is in region III there is neither. In any event these contributions to the semiclassical Green's function are known since the free space Green's function between two points x_2 and x_1 is

$$G_f(x_2, x_1, k) = -\frac{i}{4} H_0^{(+)}(kd) \approx -\frac{1}{\sqrt{8\pi kd}} \exp\{i(kd + \pi/4)\}, \quad (43.9)$$

where d is the distance between the points. For a reflection, we need to multiply by -1 and the distance is the length of the path via the reflection point. Most interesting for us, there is also a diffractive contribution to the Green's function. In equation (43.8), we recognize that the coefficient A is simply the intensity at the origin if there were no scatterer. This is therefore replaced by the Green's function to go from the source to the vertex which we label x_V . Furthermore, we recognize that $\exp(ikr)/\sqrt{kr}$ is, within a proportionality constant, the semiclassical Green's function to go from the vertex to the receiver.

Collecting these facts, we say

$$G_{\text{diff}}(x, x', k) = G_f(x, x_V, k) d(\theta, \theta') G_f(x_V, x', k), \quad (43.10)$$

where, by comparison with equations (43.8) and (43.9), we have

$$d(\theta, \theta') = \sec\left(\frac{\theta - \theta'}{2}\right) - \sec\left(\frac{\theta + \theta'}{2}\right). \quad (43.11)$$

Here θ' is the angle to the source as measured from the vertex and θ is the angle to the receiver. They were denoted as α and ϕ previously. Note that there is a symmetry between the source and receiver as we expect for a time-reversal invariant process. Also the diffraction coefficient d does not depend on which face of the half line we use to measure the angles. As we will see, a very important property of G_{diff} is that it is a simple multiplicative combination of other semiclassical Green's functions.

exercise 43.2

We now recover our deterministic perspective by realizing that we can still think of deterministic trajectories. In calculating the quantum Green's function, we sum over the contributions of various paths. These include the deterministic trajectories which connect the points and also paths which connect the points via the vertex. These have different weights as given by equations (43.9) and (43.10) but the concept of summing over deterministic paths is preserved.

For completeness, we remark that there is an exact integral representation for the Green's function in the presence of a wedge of arbitrary opening angle [3]. It can be written as

$$G(x, x', k) = g(r, r', k, \theta' - \theta) - g(r, r', k, \theta' + \theta) \quad (43.12)$$

where (r, θ) and (r', θ') are the polar coordinates of the points x and x' as measured from the vertex and the angles are measured from either face of the wedge. The function g is given by

$$g(r, r', k, \psi) = \frac{i}{8\pi\nu} \int_{C_1+C_2} d\beta \frac{H_0^+(k \sqrt{r^2 + r'^2 - 2rr' \cos \beta})}{1 - \exp(i\frac{\beta+\psi}{\nu})} \quad (43.13)$$

where $\nu = \gamma/\pi$ and γ is the opening angle of the wedge. (ie $\gamma = 2\pi$ in the case of the half plane). The contour $C_1 + C_2$ is the same as shown in figure 43.2.

The poles of this integral give contributions which can be identified with the geometric paths connecting x and x' . The saddle points at $\beta = \pm\pi$ give contributions which can be identified with the diffractive path connecting x and x' . The saddle point analysis allows us to identify the diffraction constant as

$$d(\theta, \theta') = -\frac{4 \sin \frac{\pi}{\nu}}{\nu} \frac{\sin \frac{\theta}{\nu} \sin \frac{\theta'}{\nu}}{\left(\cos \frac{\pi}{\nu} - \cos \frac{\theta+\theta'}{\nu}\right) \left(\cos \frac{\pi}{\nu} - \cos \frac{\theta-\theta'}{\nu}\right)}, \quad (43.14)$$

which reduces to (43.11) when $\nu = 2$. Note that the diffraction coefficient vanishes identically if $\nu = 1/n$ where n is any integer. This corresponds to wedge angles of $\gamma = \pi/n$ (eg. $n=1$ corresponds to a full line and $n=2$ corresponds to a right angle). This demonstration is limited by the fact that it came from a leading order asymptotic expansion but the result is quite general. For such wedge angles, we can use the method of images (we will require $2n - 1$ images in addition to the actual source point) to obtain the Green's function and there is no diffractive contribution to any order. Deterministically this corresponds to the fact that for such angles, there is no discontinuity in the dynamics. Trajectories going into the vertex can be continued out of them unambiguously. This meshes with the discussion in the introduction where we argued that diffractive effects are intimately linked with deterministic discontinuities.

The integral representation is also useful because it allows us to consider geometries such that the angles are near the optical boundaries or the wedge angle is close to π/n . For these geometries the saddle point analysis leading to (43.14) is invalid due to the existence of a nearby pole. In that event, we require a more sophisticated asymptotic analysis of the full integral representation.

Figure 43.4: The billiard considered here. The dynamics consists of free motion followed by specular reflections off the faces. The top vertex induces diffraction while the bottom one is a right angle and induces two specular geometric reflections.

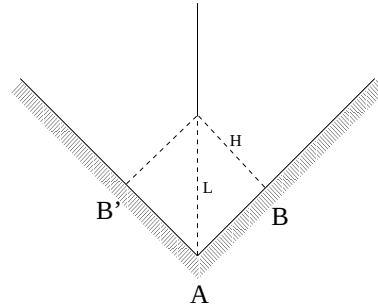
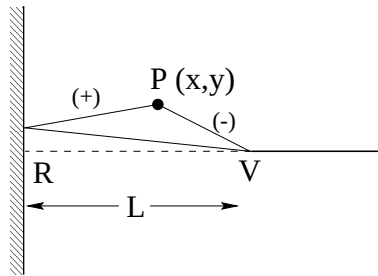


Figure 43.5: The dashed line shows a simple periodic diffractive orbit γ . Two geometric legs labeled + and - are located between vertex V and point P . The origin of the coordinate system is labelled R .



43.2 An application

Although we introduced diffraction as a correction to the purely deterministic effects, it is instructive to consider a system which can be quantized solely in terms of periodic diffractive orbits. Consider the geometry shown in figure 43.4. The deterministic mechanics consists of free motion followed by specular reflections off faces. The upper vertex is a source of diffraction while the lower one is a right angle and induces no diffraction. This is an open system, there are no bound states - only scattering resonances. However, we can still test the effectiveness of the theory in predicting them. Formally, scattering resonances are the poles of the scattering S matrix and by an identity of Balian and Bloch are also poles of the quantum Green's function. We demonstrate this fact in chapter 40 for 2-dimensional scatterers. The poles have complex wavenumber k , as for the 3-disk problem.

Let us first consider how diffractive orbits arise in evaluating the trace of G which we call $g(k)$. Specifying the trace means that we must consider all paths which close on themselves in the configuration space while stationary phase arguments for large wavenumber k extract those which are periodic - just as for deterministic trajectories. In general, $g(k)$ is given by the sum over all diffractive and geometric orbits. The contribution of the simple diffractive orbit γ (figure 43.5) to $g(k)$ is described below.

We consider a point P just a little off the path and determine the semiclassical Green's function to return to P via the vertex using (43.9) and (43.10). To leading order in y the lengths of the two geometric paths connecting P and V are $d_{\pm} = (L \pm x) + y^2 / (L \pm x)^2 / 2$ so that the phase factor $ik(d_+ + d_-)$ equals $2ikL +iky^2 / (L^2 - x^2)$.

The trace integral involves integrating over all points P and is

$$g_\gamma(k) \approx -2d_\gamma \frac{e^{i(2kL+\pi/2)}}{8\pi k} \int_0^L \frac{dx}{\sqrt{L^2-x^2}} \int_{-\infty}^{\infty} dy e^{i(ky^2 \frac{L}{L^2-x^2})}. \quad (43.15)$$

We introduced an overall negative sign to account for the reflection at the hard wall and multiplied by 2 to account for the two traversal senses, $VRPV$ and $VPRV$. In the spirit of stationary phase integrals, we have neglected the y dependence everywhere except in the exponential. The diffraction constant d_γ is the one corresponding to the diffractive periodic orbit. To evaluate the y integral, we use the identity

$$\int_{-\infty}^{\infty} d\xi e^{ia\xi^2} = e^{i\pi/4} \sqrt{\frac{\pi}{a}}, \quad (43.16)$$

and thus obtain a factor which precisely cancels the x dependence in the x integral. This leads to the rather simple result

$$g_\gamma \approx -\frac{il_\gamma}{2k} \left\{ \frac{d_\gamma}{\sqrt{8\pi k l_\gamma}} \right\} e^{i(kl_\gamma + \pi/4)} \quad (43.17)$$

where $l_\gamma = 2L$ is the length of the periodic diffractive orbit. A more sophisticated analysis of the trace integral has been done [16] using the integral representation (43.13). It is valid in the vicinity of an optical boundary and also for wedges with opening angles close to π/n .

Consider a periodic diffractive orbit with n_γ reflections off straight hard walls and μ_γ diffractions each with a diffraction constant $d_{\gamma,j}$. The total length of the orbit $L_\gamma = \sum l_{\gamma,j}$ is the sum of the various diffractive legs and l_γ is the length of the corresponding prime orbit. For such an orbit, (43.17) generalizes to

$$g_\gamma(k) = -\frac{il_\gamma}{2k} \left\{ \prod_{j=1}^{\mu_\gamma} \frac{d_{\gamma,j}}{\sqrt{8\pi k l_{\gamma,j}}} \right\} \exp \{i(kL_\gamma + n_\gamma\pi - 3\mu_\gamma\pi/4)\}. \quad (43.18)$$

Each diffraction introduces a factor of $1/\sqrt{k}$ and multi-diffractive orbits are thereby suppressed. exercise 43.3

If the orbit γ is prime then $L_\gamma = l_\gamma$. If γ is the r 'th repeat of a prime orbit β we have $L_\gamma = rl_\beta$, $n_\gamma = rp_\beta$ and $\mu_\gamma = r\sigma_\beta$, where l_β , p_β and σ_β all refer to the prime orbit. We can then write

$$g_\gamma = g_{\beta,r} = -\frac{il_\beta}{2k} t_\beta^r \quad (43.19)$$

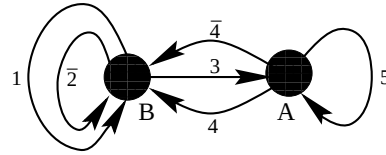
where

$$t_\beta = \left\{ \prod_{j=1}^{\sigma_\beta} \frac{d_{\beta,j}}{\sqrt{8\pi k l_{\beta,j}}} \right\} \exp \{i(kl_\beta + p_\beta\pi - 3\sigma_\beta\pi/4)\}. \quad (43.20)$$

It then makes sense to organize the sum over diffractive orbits as a sum over the prime diffractive orbits and a sum over the repetitions

$$g_{\text{diff}}(k) = \sum_{\beta} \sum_{r=1}^{\infty} g_{\beta,r} = -\frac{i}{2k} \sum_{\beta} l_\beta \frac{t_\beta}{1-t_\beta}. \quad (43.21)$$

Figure 43.6: The two-node transition graph with all the diffractive processes connecting the nodes.



We cast this as a logarithmic derivative (22.6) by noting that $\frac{dt_\beta}{dk} = i l_\beta t_\beta - \sigma_\beta t_\beta / 2k$ and recognizing that the first term dominates in the semiclassical limit. It follows that

$$g_{\text{diff}}(k) \approx \frac{1}{2k} \frac{d}{dk} \left\{ \ln \prod_{\beta} (1 - t_\beta) \right\}. \quad (43.22)$$

In the case that there are only diffractive periodic orbits - as in the geometry of figure 43.4 - the poles of $g(k)$ are the zeros of a dynamical zeta function

$$1/\zeta(k) = \prod_{\beta} (1 - t_\beta). \quad (43.23)$$

For geometric orbits, this function would be evaluated with a cycle expansion as discussed in chapter 23. However, here we can use the multiplicative nature of the weights t_β to find a closed form representation of the function using a transition graph, as in chapter 17. This multiplicative property of the weights follows from the fact that the diffractive Green's function (43.10) is multiplicative in segment semiclassical Green's functions, unlike the geometric case.

There is a reflection symmetry in the problem which means that all resonances can be classified as even or odd. Because of this, the dynamical zeta function factorizes as $1/\zeta = 1/\zeta_+ \zeta_-$ (as explained in example 25.9) and we determine $1/\zeta_+$ and $1/\zeta_-$ separately using the ideas of symmetry decomposition of chapter 25.

In the transition graph shown in figure 43.6, we enumerate all processes. We start by identifying the fundamental domain as just the right half of figure 43.4. There are two nodes which we call A and B . To get to another node from B , we can diffract (always via the vertex) in one of three directions. We can diffract back to B which we denote as process 1. We can diffract to B 's image point B' and then follow this by a reflection. This process we denote as $\bar{2}$ where the bar indicates that it involves a reflection. Third, we can diffract to node A . Starting at A we can also diffract to a node in three ways. We can diffract to B which we denote as 4. We can diffract to B' followed by a reflection which we denote as $\bar{4}$. Finally, we can diffract back to A which we denote as process 5. Each of these processes has its own weight which we can determine from the earlier discussion. First though, we construct the dynamical zeta functions.

The dynamical zeta functions are determined by enumerating all closed loops which do not intersect themselves in figure 43.6. We do it first for $1/\zeta_+$ because that is simpler. In that case, the processes with bars are treated on an equal footing as the others. Appealing back to sect. 25.5 we find

$$\begin{aligned} 1/\zeta_+ &= 1 - t_1 - t_2 - t_5 - t_3 t_4 - t_3 t_{\bar{4}} + t_5 t_1 + t_5 t_2, \\ &= 1 - (t_1 + t_2 + t_5) - 2t_3 t_4 + t_5(t_1 + t_2) \end{aligned} \quad (43.24)$$

where we have used the fact that $t_4 = t_{\bar{4}}$ by symmetry. The last term has a positive sign because it involves the product of shorter closed loops. To calculate $1/\zeta_-$, we note that the processes with bars have a relative negative sign due to the group theoretic weight. Furthermore, process 5 is a boundary orbit (see sect. 25.4.3) and only affects the even resonances - the terms involving t_5 are absent from $1/\zeta_-$. The result is

$$\begin{aligned} 1/\zeta_- &= 1 - t_1 + t_2 - t_3 t_4 + t_3 t_{\bar{4}}, \\ &= 1 - (t_1 - t_2). \end{aligned} \quad (43.25)$$

Note that these expressions have a finite number of terms and are not in the form of a curvature expansion, as for the 3-disk problem.

exercise 43.4

It now just remains to fix the weights. We use equation (43.20) but note that each weight involves just one diffraction constant. It is then convenient to define the quantities

$$u_A^2 = \frac{\exp\{i(2kL + 2\pi)\}}{\sqrt{16\pi kL}} \quad u_B^2 = \frac{\exp\{i(2kH + \pi)\}}{\sqrt{16\pi kH}}. \quad (43.26)$$

The lengths L and $H = L/\sqrt{2}$ are defined in figure 43.4; we set $L = 1$ throughout. Bouncing inside the right angle at A corresponds to two specular reflections so that $p = 2$. We therefore explicitly include the factor $\exp(i2\pi)$ in (43.26) although it is trivially equal to one. Similarly, there is one specular reflection at point B giving $p = 1$ and therefore a factor of $\exp(i\pi)$. We have defined u_A and u_B because, together with some diffraction constants, they can be used to construct all of the weights. Altogether we define four diffraction coefficients: d_{AB} is the constant corresponding to diffracting from B to A and is found from (43.11) with $\theta' = 3\pi/4$ and $\theta = \pi$ and equals $2 \sec(\pi/8) \approx 2.165$. With analogous notation, we have d_{AA} and $d_{BB} = d_{B'B}$ which equal 2 and $1 + \sqrt{2}$ respectively. $d_{ij} = d_{ji}$ due to the Green's function symmetry between source and receiver referred to earlier. Finally, there is the diffractive phase factor $s = \exp(-i3\pi/4)$ each time there is a diffraction. The weights are then as follows:

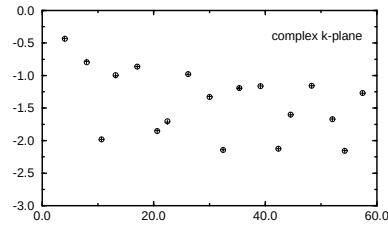
$$\begin{aligned} t_1 &= s d_{BB} u_B^2 & t_2 &= s d_{B'B} u_B^2 & t_3 &= t_4 = t_{\bar{4}} = s d_{AB} u_A u_B \\ t_5 &= s d_{AA} u_A^2. \end{aligned} \quad (43.27)$$

Each weight involves two u 's and one d . The u 's represent the contribution to the weight from the paths connecting the nodes to the vertex and the d gives the diffraction constant connecting the two paths.

The equality of d_{BB} and $d_{B'B}$ implies that $t_1 = t_2$. From (43.25) this means that there are no odd resonances because 1 can never equal 0. For the even resonances equation (43.24) is an implicit equation for k which has zeros shown in figure 43.7.

For comparison we also show the result from an exact quantum calculation. The agreement is very good right down to the ground state - as is so often the

Figure 43.7: The even resonances of the wedge scatterer of figure 43.4 plotted in the complex k -plane, with $L = 1$. The exact resonances are represented as circles and their semiclassical approximations as crosses.



case with semiclassical calculations. In addition we can use our dynamical zeta function to find arbitrarily high resonances and the results actually improve in that limit. In the same limit, the exact numerical solution becomes more difficult to find so the dynamical zeta function approximation is particularly useful in that case.

exercise 43.5

In general a system will consist of both geometric and diffractive orbits. In that case, the full dynamical zeta function is the product of the geometric zeta function and the diffractive one. The diffractive weights are typically smaller by order $O(1/\sqrt{k})$ but for small k they can be numerically competitive so that there is a significant diffractive effect on the low-lying spectrum. It might be expected that higher in the spectrum, the effect of diffraction is weaker due to the decreasing weights. However, it should be pointed out that an analysis of the situation for creeping diffraction [11] concluded that the diffraction is actually *more* important higher in the spectrum due to the fact that an ever greater fraction of the orbits need to be corrected for diffractive effects. The equivalent analysis has not been done for edge diffraction but a similar conclusion can probably be expected.

To conclude this chapter, we return to the opening paragraph and discuss the possibility of doing such an analysis for helium. The important point which allowed us to successfully analyze the geometry of figure 43.4 is that when a trajectory is near the vertex, we can extract its diffraction constant without reference to the other facets of the problem. We say, therefore, that this is a “local” analysis for the purposes of which we have “turned off” the other aspects of the problem, namely sides AB and AB' . By analogy, for helium, we would look for some simpler description of the problem which applies near the three body collision. However, there is nothing to “turn off.” The local problem is just as difficult as the global one since they are precisely the same problem, just related by scaling. Therefore, it is not at all clear that such an analysis is possible for helium.

Résumé

In this chapter we have discovered new types of periodic orbits contributing to the semiclassical traces and determinants. Unlike the periodic orbits we had seen so far, these are not true deterministic orbits. They are generated by singularities of the scattering potential. In these singular points the deterministic dynamics has no unique definition, and the deterministic orbits hitting the singularities can be continued in many different directions. While the deterministic mechanics does not know which way to go, quantum mechanics solves the dilemma by allowing

us to continue in all possible directions. The likelihoods of different paths are given by the quantum mechanical weights called diffraction constants. The total contribution to a trace from such orbit is given by the product of transmission amplitudes between singularities and diffraction constants of singularities. The weights of diffractive periodic orbits are at least of order $1/\sqrt{k}$ weaker than the weights associated with deterministically realizable orbits, and their contribution at large energies is therefore negligible. Nevertheless, they can strongly influence the low lying resonances or energy levels. In some systems, such as the N disk scattering the diffraction effects do not only perturb semiclassical resonances, but can also create new low energy resonances. Therefore it is always important to include the contributions of diffractive periodic orbits when semiclassical methods are applied at low energies.

Commentary

Remark 43.1. Deterministic discontinuities. Various classes of discontinuities for billiard and potential problems discussed in the literature:

- a grazing condition such that some trajectories hit a smooth surface while others are unaffected, refs. [10, 11, 20, 23]
- a vertex such that trajectories to one side bounce differently from those to the other side, refs. [2, 9, 20–22].
- a point scatterer [14, 15] or a magnetic flux line [1, 13] such that we do not know how to continue deterministic mechanics through the discontinuities.

Remark 43.2. Geometrical theory of diffraction. In the above discussion we borrowed heavily from the ideas of Keller who was interested in extending the geometrical ray picture of optics to cases where there is a discontinuity. He maintained that we could hang onto that ray-tracing picture by allowing rays to strike the vertex and then leave at any angle with amplitude (43.8). Both he and Sommerfeld were thinking of optics and not quantum mechanics and they did not phrase the results in terms of semiclassical Green's functions but the essential idea is the same.

Remark 43.3. Generalizations. Consider the effect of replacing our half line by a wedge of angle γ_1 and the right angle by an arbitrary angle γ_2 . If $\gamma_2 > \gamma_1$ and $\gamma_2 \geq \pi/2$ this is an open problem whose solution is given by equations (43.24) and (43.25) (there will then be odd resonances) but with modified weights reflecting the changed geometry [22]. (For $\gamma_2 < \pi/2$, more diffractive periodic orbits appear and the dynamical zeta functions are more complicated but can be calculated with the same machinery.) When $\gamma_2 = \gamma_1$, the problem in fact has bound states [8, 24]. This last case has been of interest in studying electron transport in mesoscopic devices and in microwave waveguides. However we can not use our formalism as it stands because the diffractive periodic orbits for this geometry lie right on the border between illuminated and shadowed regions so that equation (43.7) is invalid. Even the more uniform derivation of ref. [16] fails for that particular geometry, the problem being that the diffractive orbit actually lives on the edge of a family of geometric orbits and this makes the analysis still more difficult.

Remark 43.4. Diffraction Green's functions. The result (43.17) is proportional to the length of the orbit times the semiclassical Green's function (43.9) to go from the vertex back to itself along the deterministic path. The multi-diffractive formula (43.18) is proportional to the total length of the orbit times the product of the semiclassical Green's functions to go from one vertex to the next along deterministic paths. This result generalizes to any system — either a pinball or a potential — which contains point singularities such that we can define a diffraction constant as above. The contribution to the trace of the semiclassical Green's function coming from a diffractive orbit which hits the singularities is proportional to the total length (or period) of the orbit times the product of semiclassical Green's functions in going from one singularity to the next. This result first appeared in reference [20] and a derivation can be found in ref. [2]. A similar structure also exists for creeping [20].

Remark 43.5. Diffractive orbits for hydrogenic atoms. An analysis in terms of diffractive orbits has been made in a different atomic physics system, the response of hydrogenic atoms to strong magnetic fields [5, 6]. In these systems, a single electron is highly excited and takes long traversals far from the nucleus. Upon returning to a hydrogen nucleus, it is re-ejected with the reversed momentum as discussed in chapter 42. However, if the atom is not hydrogen but sodium or some other atom with one valence electron, the returning electron feels the charge distribution of the core electrons and not just the charge of the nucleus. This so-called quantum defect induces scattering in addition to the deterministic re-ejection present in the hydrogen atom. (In this case the local analysis consists of neglecting the magnetic field when the trajectory is near the nucleus.) This is formally similar to the vertex which causes both specular reflection and diffraction. There is then additional structure in the Fourier transform of the quantum spectrum corresponding to the induced diffractive orbits, and this has been observed experimentally [4, 7, 12].

References

- [1] M. Brack, R. K. Bhaduri, J. Law, C. Maier, and M. V. N. Murthy, “Effect of a magnetic flux line on the quantum beats in the Hénon-Heiles level density”, *Chaos* **5**, 317–329 (1995).
- [2] H. Bruus and N. D. Whelan, “Edge diffraction, trace formulae and the cardioid billiard”, *Nonlinearity* **9**, 1023–1047 (1996).
- [3] H. S. Carslaw, “Diffraction of waves by a wedge of any angle”, *Proc. London Math. Soc. (Ser. 2)* **18**, 291–306 (1920).
- [4] M. Courtney, H. Jiao, N. Spellmeyer, and D. Kleppner, “Long-period orbits in the Stark spectrum of lithium”, *Phys. Rev. Lett.* **73**, 1340–1343 (1994).
- [5] P. A. Dando, T. S. Monteiro, D. Delande, and K. T. Taylor, “Beyond periodic orbits: An example in nonhydrogenic atoms”, *Phys. Rev. Lett.* **74**, 1099–1102 (1995).
- [6] P. A. Dando, T. S. Monteiro, and S. M. Owen, “Periodic orbit theory for Rydberg atoms in external fields”, *Phys. Rev. Lett.* **80**, 2797–2800 (1998).

- [7] D. Delande, K. T. Taylor, M. H. Halley, T. van der Veldt, W. Vassen, and W. Hogervorst, “Scaled energy spectra of non-hydrogenic Rydberg atoms in a magnetic field”, *J. Phys. B* **27**, 2771 (1994).
- [8] P. Exner, P. Šeba, and P. Št’oviček, “On existence of a bound state in an L-shaped waveguide”, *Czech. J. Phys. B* **39**, 1181–1191 (1989).
- [9] N. Pavloff and C. Schmit, “Diffractive orbits in quantum billiards”, *Phys. Rev. Lett.* **75**, 61–64 (1995).
- [10] H. Primack, H. Schanz, U. Smilansky, and I. Ussishkin, “Penumbra diffraction in the quantization of dispersing billiards”, *Phys. Rev. Lett.* **76**, 1615–1618 (1996).
- [11] H. Primack, H. Schanz, U. Smilansky, and I. Ussishkin, “Penumbra diffraction in the semiclassical quantization of concave billiards”, *J. Phys. A* **30**, 6693 (1997).
- [12] G. Raithel, H. Held, L. Marmet, and H. Walther, “Rubidium Rydberg atoms in strong static fields”, *J. Phys. B* **27**, 2849 (1994).
- [13] S. M. Reimann, M. Brack, A. G. Magner, J. Blaschke, and M. V. N. Murthy, “Circular quantum billiard with a singular magnetic flux line”, *Phys. Rev. A* **53**, 39–48 (1996).
- [14] P. E. Rosenqvist, N. D. Whelan, and A. Wirzba, “Small disks and semiclassical resonances”, *J. Phys. A* **29**, 5441–5453 (1996).
- [15] P. Šeba, “Wave chaos in singular quantum billiard”, *Phys. Rev. Lett.* **64**, 1855–1858 (1990).
- [16] M. Sieber, N. Pavloff, and C. Schmit, “Uniform approximation for diffractive contributions to the trace formula in billiard systems”, *Phys. Rev. E* **55**, 2279–2299 (1997).
- [17] A. Sommerfeld, “Mathematische Theorie der Diffraction”, *Mathem. Ann.* **47**, 317–374 (1896).
- [18] A. Sommerfeld, *Optics*, Vol. 4, Lectures on Theoretical Physics (Academic Press, 1954).
- [19] A. Sommerfeld, *Mathematical Theory of Diffraction* (Birkhäuser, Boston, MA, 2004).
- [20] G. Vattay, A. Wirzba, and P. E. Rosenqvist, “Periodic orbit theory of diffraction”, *Phys. Rev. Lett.* **73**, 2304–2307 (1994).
- [21] N. D. Whelan, “Geometric and diffractive orbits in the scattering from confocal hyperbolas”, *Phys. Rev. E* **51**, 3778–3781 (1995).
- [22] N. D. Whelan, “Semiclassical quantization using diffractive orbits”, *Phys. Rev. Lett.* **76**, 2605–2608 (1996).
- [23] A. Wirzba, “Validity of the semiclassical periodic orbit approximation in the two- and three-disk problems”, *Chaos* **2**, 77–83 (1992).
- [24] H. Wu, D. W. L. Sprung, and J. Martorell, “Electronic properties of a quantum wire with arbitrary bending angle”, *J. Appl. Phys.* **72**, 151–154 (1992).

43.3 Exercises

- 43.1. **Stationary phase integral.** Evaluate the two stationary phase integrals corresponding to contours E_1 and E_2 of figure 43.3 and thereby verify (43.7).

(N. Whelan)

- 43.2. **Scattering from a small disk** Imagine that instead of a wedge, we have a disk whose radius a is much smaller than the typical wavelengths we are considering. In that limit, solve the quantum scattering problem - find the scattered wave which result from an incident plane wave. You can do this by the method of partial waves - the analogous three dimensional problem is discussed in most quantum textbooks. You should find that only the $m = 0$ partial wave contributes for small a . Following the discussion above, show that the diffraction constant is

$$d = \frac{2\pi}{\log\left(\frac{2}{ka}\right) - \gamma_e + i\frac{\pi}{2}} \quad (43.28)$$

where $\gamma_e = 0.577 \dots$ is Euler's constant. Note that in this limit d depends weakly on k but not on the scattering angle.

(N. Whelan)

- 43.3. **Several diffractive legs.** Derive equation (43.18). The calculation involves considering slight variations of the diffractive orbit as in the simple case discussed above. Here it is more complicated because there are more diffractive arcs - however you should convince yourself that a slight variation of the diffractive orbit only affects one leg at a time.

(N. Whelan)

- 43.4. **Unsymmetrized dynamical zeta function.** Assume you know nothing about symmetry decomposition. Construct the 3-node transition graph for figure 43.1 by considering A , B and B' to be physically distinct. Write down the corresponding dynamical zeta function and check explicitly that for $B = B'$ it factorizes into the product of the even and odd dynamical zeta functions. Why is there no term t_2 in the full dynamical zeta function?

(N. Whelan)

- 43.5. **Three point scatterers.**

Consider the limiting case of the three disk game of pinball of figure 1.1 where the disks are very much smaller than their spacing R . Use the results of exercise 43.2 to construct the desymmetrized dynamical zeta functions, as in sect. 25.6. You should find $1/\zeta_{A_1} = 1 - 2t$ where $t = de^{i(kR-3\pi/4)} / \sqrt{8\pi kR}$. Compare this formula with that from chapter 14. By assuming that the real part of k is much greater than the imaginary part show that the positions of the resonances are $k_n R = \alpha_n - i\beta_n$ where $\alpha_n = 2\pi n + 3\pi/4$, $\beta_n = \log(\sqrt{2\pi\alpha_n}/d)$ and n is a non-negative integer. (See also ref. [14].)

(N. Whelan)

Epilogue

Nowadays, whatever the truth of the matter may be (and we will probably never know), the simplest solution is no longer emotionally satisfying. Everything we know about the world militates against it. The concepts of indeterminacy and chaos have filtered down to us from the higher sciences to confirm our nagging suspicions.

—L. Sante, “Review of ‘American Tabloid’ by James Ellroy,” *New York Review of Books* (May 11, 1995)

A MOTION on a strange attractor can be approximated by shadowing long orbits by sequences of nearby shorter periodic orbits. This notion has here been made precise by approximating orbits by prime cycles, and evaluating associated curvatures. A curvature measures the deviation of a long cycle from its approximation by shorter cycles; the smoothness of the dynamical system implies exponential fall-off for (almost) all curvatures. We propose that the theoretical and experimental non-wandering sets be expressed in terms of the symbol sequences of short cycles (a topological characterization of the spatial layout of the non-wandering set) and their eigenvalues (metric structure)

Cycles as the skeleton of chaos

We wind down this all-too-long treatise by asking: why cycle?

We tend to think of a dynamical system as a smooth system whose evolution can be followed by integrating a set of differential equations. Traditionally one used integrable motions as zeroth-order approximations to physical systems, and accounted for weak nonlinearities perturbatively. However, when the evolution is actually followed through to asymptotic times, one discovers that the strongly nonlinear systems show an amazingly rich structure which is not at all apparent in their formulation in terms of differential equations. In particular, the periodic orbits are important because they form the *skeleton* onto which all trajectories trapped for long times cling. This was already appreciated century ago by H. Poincaré, who, describing in *Les méthodes nouvelles de la mécanique céleste* his discovery of homoclinic tangles, mused that “the complexity of this figure will be striking, and I shall not even try to draw it.” Today such drawings are cheap and

plentiful; but Poincaré went a step further and, noting that hidden in this apparent chaos is a rigid skeleton, a tree of *cycles* (periodic orbits) of increasing periods and self-similar structure, suggested that the cycles should be the key to chaotic dynamics.

The zeroth-order approximations to harshly chaotic dynamics are very different from those for the nearly integrable systems: a good starting approximation here is the stretching and kneading of a baker's map, rather than the winding of a harmonic oscillator.

For low dimensional deterministic dynamical systems description in terms of cycles has many virtues:

1. cycle symbol sequences are *topological* invariants: they give the spatial layout of a non-wandering set
2. cycle eigenvalues are *metric* invariants: they give the scale of each piece of a non-wandering set
3. cycles are *dense* on the asymptotic non-wandering set
4. cycles are ordered *hierarchically*: short cycles give good approximations to a non-wandering set, longer cycles only refinements. Errors due to neglecting long cycles can be bounded, and typically fall off exponentially or super-exponentially with the cutoff cycle length
5. cycles are *structurally robust*: for smooth flows eigenvalues of short cycles vary slowly with smooth parameter changes
6. asymptotic averages (such as correlations, escape rates, quantum mechanical eigenstates and other “thermodynamic” averages) can be efficiently computed from short cycles by means of *cycle expansions*

Points 1, 2: That the cycle topology and eigenvalues are invariant properties of dynamical systems follows from elementary considerations. If the same dynamics is given by a map f in one set of coordinates, and a map g in the next, then f and g (or any other good representation) are related by a reparametrization and a coordinate transformation $f = h^{-1} \circ g \circ h$. As both f and g are arbitrary representations of the dynamical system, the explicit form of the conjugacy h is of no interest, only the properties invariant under any transformation h are of general import. The most obvious invariant properties are topological; a fixed point must be a fixed point in any representation, a trajectory which exactly returns to the initial point (a cycle) must do so in any representation. Furthermore, a good representation should not mutilate the data; h must be a smooth transformation which maps nearby periodic points of f into nearby periodic points of g . This smoothness guarantees that the cycles are not only topological invariants, but that their linearized neighborhoods are also metrically invariant. In particular, the cycle eigenvalues (eigenvalues of the Jacobian matrix $df^n(x)/dx$ of periodic orbits $f^n(x) = x$) are invariant.

Point 5: An important virtue of cycles is their *structural robustness*. Many quantities customarily associated with dynamical systems depend on the notion of “structural stability,” i.e., robustness of non-wandering set to small parameter variations.

Still, the sufficiently short unstable cycles are structurally robust in the sense that they are only slightly distorted by such parameter changes, and averages computed using them as a skeleton are insensitive to small deformations of the non-wandering set. In contrast, lack of structural stability wreaks havoc with long time averages such as Lyapunov exponents, for which there is no guarantee that they converge to the correct asymptotic value in any finite time numerical computation.

The main recent theoretical advance is **point 4:** we now know how to control the errors due to neglecting longer cycles. As we seen above, even though the number of invariants is infinite (unlike, for example, the number of Casimir invariants for a compact Lie group) the dynamics can be well approximated to any finite accuracy by a small finite set of invariants. The origin of this convergence is geometrical, see appendix 28, and for smooth flows the convergence of cycle expansions can even be super-exponential.

The cycle expansions such as (23.8) outperform the pedestrian methods such as extrapolations from the finite cover sums (27.2) for a number of reasons. The cycle expansion is a better averaging procedure than the naive box counting algorithms because the strange attractor is here pieced together in a topologically invariant way from neighborhoods (“space average”) rather than explored by a long ergodic trajectory (“time average”). The cycle expansion is co-ordinate and reparametrization invariant - a finite n th level sum (27.2) is not. Cycles are of finite period but infinite duration, so the cycle eigenvalues are already evaluated in the $n \rightarrow \infty$ limit, but for the sum (27.2) the limit has to be estimated by numerical extrapolations. And, crucially, the higher terms in the cycle expansion (23.8) are deviations of longer prime cycles from their approximations by shorter cycles. Such combinations vanish exactly in piecewise linear approximations and fall off exponentially for smooth dynamical flows.

In the above we have reviewed the general properties of the cycle expansions; those have been applied to a series of examples of low-dimensional chaos: 1-d strange attractors, the period-doubling repeller, the Hénon-type maps and the mode locking intervals for circle maps. The cycle expansions have also been applied to the irrational windings set of critical circle maps, to the Hamiltonian period-doubling repeller, to a Hamiltonian three-disk game of pinball, to the three-disk quantum scattering resonances and to the extraction of correlation exponents. Feasibility of analysis of experimental non-wandering set in terms of cycles is discussed in ref. [1].

Homework assignment

“Lo! thy dread empire Chaos is restor’d, Light dies before
thy uncreating word; Thy hand, great Anarch, lets the cur-
tain fall, And universal darkness buries all.”

—Alexander Pope, *The Dunciad*

We conclude cautiously with a homework assignment posed May 22, 1990 (the original due date was May 22, 2000, but alas...):

1. **Topology** Develop optimal sequences (“continued fraction approximants”) of finite subshift approximations to generic dynamical systems. Apply to (a) the Hénon map, (b) the Lorenz flow and (c) the Hamiltonian standard map.
2. **Nonhyperbolicity** Incorporate power-law (marginal stability orbits, “intermittency”) corrections into cycle expansions. Apply to long-time tails in the Hamiltonian diffusion problem.
3. **Phenomenology** Carry through a convincing analysis of a genuine experimentally extracted data set in terms of periodic orbits.
4. **Invariants** Prove that the scaling functions, or the cycles, or the spectrum of a transfer operator are the maximal set of invariants of an (physically interesting) dynamically generated non-wandering set.
5. **Field theory** Develop a periodic orbit theory of systems with many unstable degrees of freedom. Apply to (a) coupled lattices, (b) cellular automata, (c) neural networks.
6. **Tunneling** Add complex time orbits to quantum mechanical cycle expansions (WKB theory for chaotic systems).
7. **Unitarity** Evaluate corrections to the Gutzwiller semiclassical periodic orbit sums. (a) Show that the zeros (energy eigenvalues) of the appropriate Selberg products are real. (b) Find physically realistic systems for which the “semiclassical” periodic orbit expansions yield the exact quantization.
8. **Atomic spectra** Compute the helium spectrum from periodic orbit expansions (already accomplished by Wintgen and Tanner!).
9. **Symmetries** Include fermions, gauge fields into the periodic orbit theory.
10. **Quantum field theory** Develop quantum theory of systems with infinitely many deterministically unstable degrees of freedom. Apply to (a) quark confinement (b) early universe (c) the brain.

Conclusion

Good-bye. I am leaving because I am bored.

—George Saunders' dying words

Nadie puede escribir un libro. Para Que un libro sea verdaderamente, Se requieren la aurora y el poniente Siglos, armas y el mar que une y separa.

—Jorge Luis Borges El Hacedor, *Ariosto y los arabes*

The buttler did it.

Contributors

No man but a blockhead ever wrote except for money
—Samuel Johnson

This book is a result of collaborative labors of many people over a span of several decades. Coauthors of a chapter or a section are indicated in the byline to the chapter/section title. If you are referring to a specific coauthored section rather than the entire book, cite it as (for example):

C. Chandre, F.K. Diakonov and P. Schmelcher, section “Discrete cyclist relaxation method,” in P. Cvitanović, R. Artuso, R. Mainieri, G. Tanner and G. Vattay, *Chaos: Classical and Quantum* (Niels Bohr Institute, Copenhagen 2010); ChaosBook.org/version13.

Do not cite chapters by their numbers, as those change from version to version. Chapters without a byline are written by Predrag Cvitanović. Friends whose contributions and ideas were invaluable to us but have not contributed written text to this book, are credited in the acknowledgments.

Roberto Artuso

19 Transporting densities	381
21.2 A trace formula for flows	423
27.3 Correlation functions	547
29 Intermittency	580
24 Deterministic diffusion	475
Appendix A31.2: Implementing evolution	1059

Ronnie Mainieri

2 Flows	39
3.2 The Poincaré section of a flow	74
4 Local stability	86
A2.1 Understanding flows	919
14.1 Temporal ordering: itineraries	261
Appendix A1: A brief history of chaos	872

Gábor Vattay

Gregor Tanner

29 Intermittency	580
Appendix A8.2: Jacobians of Hamiltonian flows	948

Arindam Basu

Rössler flow figures, tables, cycles in chapters 14, 16 and exercise 7.1

Ofer Biham

34.1 Cyclists relaxation method	696

Daniel Borrero Oct 23 2008, soluCycles.tex

Solution 16.1

N. Burak Budanur

13.5 First Fourier mode slice	242
13.2 In-slice stability	256
Solution 12.5	
Solution 12.6	

Cristel Chandre

34.1 Cyclists relaxation method	696
34.2 Discrete cyclists relaxation methods	699

Freddy Christiansen

7.1 One-dimensional mappings	137
16.2 Multipoint shooting method	324

Per Dahlqvist

29 Intermittency	580
34.3 Orbit length extremization method for billiards	701

Carl P. Dettmann

23.7 Stability ordering of cycle expansions	462

Fotis K. Diakonos

34.2 Discrete cyclists relaxation methods	699

G. Bard Ermentrout

Exercise 5.1

Mitchell J. Feigenbaum

Appendix A8.1: Symplectic invariance	947

Sarah Flynn

solutions 3.5 and 3.6

Matjaž Gomilšek

17.3 Transition graphs: stroll from link to link	338

Jonathan Halcrow

Example 3.3: Sections of Lorenz flow	81
Example 4.6: Stability of Lorenz flow equilibria	104
Example 4.7: Lorenz flow: Global portrait	105

Example 11.8: Desymmetrization of Lorenz flow	206
Example 11.9: Lorenz flow in doubled-polar angle representation ...	207
Example 14.4: Lorenz flow: a 1-dimensional return map	282
Exercise 11.5 and figure 2.5	

Kai T. Hansen

14.3 Unimodal map symbolic dynamics	266
18.5 Topological zeta function for an infinite partition	358
14.5 Kneading theory	270
figures throughout the text	

Rainer Klages

Figure 24.6

Yueheng Lan

Solutions 1.1, 2.2, 2.3, 2.4, 2.5, 11.3, 15.7, 14.6, 19.1, 19.2, 19.3, 19.5,
19.7, 19.10, 6.3 and figures 1.8, 11.4, 11.5 14.5,

Bo Li

Solutions 36.2, 36.1, 37.2

Han Liang

Example 14.12 Adler-Weiss partition of the cat map state space	286
Example 18.13 A simple stable/unstable manifolds pair	373
Example 18.14 Cat map period-4 prime cycles	374
Figure A24.1, A24.2	

Norman Lebovitz

Example 15.1 A simple stable/unstable manifolds pair	314

Joachim Mathiesen

6.2 Lyapunov exponents	126
Rössler flow figures, tables, cycles in sect. 6.2 and exercise 7.1	

Yamato Matsuoka

Figure 15.4

Radford Mitchell, Jr.

Example 3.4

Rytis Paškauskas

4.6 Stability of return maps	93
5.5 Stability of return map cycles	117
Exercises 2.8, 3.1, 4.4 and solution 4.1	

Adam Prügel-Bennet

Solutions 1.2, 2.11, 9.1, 20.2, 23.2 28.3, 34.1.

Lamberto Rondoni

19 Transporting densities	381

16.1.1 Cycles from long time series	322
27.2.1 Unstable periodic orbits are dense	546
Table 18.2	

Juri Rolf

Solution 28.3

Per E. Rosenqvist

Exercises, figures throughout the text

Hans Henrik Rugh

28 Why does it work?	555

Luis Saldana

solution 10.3

Peter Schmelcher

34.2 Discrete cyclists relaxation methods	699

Evangelos Siminos

Example 3.3: Sections of Lorenz flow	81
Example 4.6: Stability of Lorenz flow equilibria	104
Example 4.7: Lorenz flow: Global portrait	105
Example 11.8: Desymmetrization of Lorenz flow	206
Example 11.9: Lorenz flow in doubled-polar angle representation ...	207
Example 14.4: Lorenz flow: a 1-dimensional return map	282
Exercise 11.5	
Solution 13.1	

Gábor Simon

Rössler flow figures, tables, cycles in chapters 2, 16 and exercise 7.1

Sara A. Solla

Example A20.4: Unbiased sample variance	1002
Example A20.5: Standard error of the mean	1003
Exercises 2.13, A20.2 and A20.3	

Edward A. Spiegel

2 Flows	39
19 Transporting densities	381

Luz V. Vela-Arevalo

8.1 Hamiltonian flows	147
Exercises 8.1, 8.3, 8.5	

Lei Zhang

Solutions 1.1, 2.1

Index

- abscissa of absolute convergence, 464
- accelerator mode, 479
- action, 415, 726, 739, 748
 - free, 198
 - helium, 804
 - relation to period, 811
- adjacency matrix, 336, 369, *see* transition matrix
- admissible
 - periodic points, 352
 - sequence, 307
 - trajectories, number of, 350
- Airy
 - equation, 731
 - function, 691, 731, 733, 792, 793, 795, 800
 - at a bifurcation, 691
 - integral, 731
- algebra, 924
 - associative, 924
 - Lie, 924
- allowable itinerary, *see* admissible
- alphabet, 262
- alternating binary tree, 276, 286
- alternative periodic orbit theories, 465
- analyticity
 - domain, 433
- angle doubling map, 277
- anisotropic
 - diffusion, 673
 - Kepler potential, 201, 467, 1055
- anomalous diffusion, 483
- Anosov
 - flows, 308
- anti-hermitian
 - generator, 230, 231
- antiharmonic extension, 986
- arc, 338, *see* edge
- area preserving
 - Hénon map, 162
 - map, 961
- Arnoldi iteration, 99
- Artin-Mazur zeta function, 357
- associative algebra, 924
- Atiyah-Singer
 - index theory, 1011
- atlas, 305
- attractor
 - basin, 44
 - Hénon, 135
 - strange, 31, 44, 63, 124
- Aubry-Mather theory, 705
- AUTO, 59
- autonomous flow, 46
- average
 - Birkhoff, 389, 403, 414
 - chaotic, 592
 - space, 388, 404
 - time, 389, 414
- averaging, 29
- axiom A, 465, 567, 571, 573, 943
- baker's map, 173, 298
- Balmer spectrum, 713
- basin of attraction, 44
- basis
 - standard, 923
 - vector, 923
- BER
 - approximation, 603
- Bernoulli, 873
 - polynomials, 575
 - shift, 277, 284, 483, 554, 563, 572, 575, 579, 584, 606, 876, 878, 971, 974, 1089
 - shift eigenfunctions, 578
 - shift return times, 606
- Berry-Keating conjecture, 898
- Bessel function, 784
 - addition theorem, 786
- bi-infinite itinerary, 273
- bifurcation
 - Airy function approximation, 691
 - bizarre, 985
 - generic, 173
 - Hopf, 649
 - period doubling, 969
 - saddle point, 970
 - saddle-node, 79
 - sequence
 - unimodal map, 969
- billiard, 168–172

- map, 169
- stability, 122, 170
- stadium, 168, 173, 176, 201, 607, 1053, 1055, 1073
- binary
 - prime cycles, 303, 315, 369
 - symbolic dynamics
 - collinear helium, 807
 - tree, alternating, 276, 286
- Birkhoff
 - average, 389, 403, 414
 - coordinates, 169, 173, 177, 317
 - ergodic theorem, 389
 - normal form, 1107, 1116
 - Poincaré-Birkhoff theorem, 164
 - sum, 403, 404, 408, 414, 417, 422, 437, 450
- Birkhoff average, 617
- Birkhoff, G. D., 414
- bit shift map, 277
- block
 - finite sequence, 273
- block, pruned, 275
- Bohr
 - helium, 803, 813
 - Uetli Schwur, 891
- Bohr-de Broglie picture, 713
- Bohr-Sommerfeld quantization, 713, 733, 764, 892
- Boltzmann
 - equation, 25, 486
 - factor, 1076, 1089
 - stosszahlansatz, 25
- Boltzmann, L., 25, 875
- boredom, 844, 1067
- Borges, J.L., 844
- Botox, 69
- boundary orbits, 198, 509
- bounded operators, 1123
- Bourbaki, N., 83
- Bowen, R., 31
- brain, rat, 4, 33
- branch cut, 588
 - singularity, 589
- Brownian noise, 675
- Bunimovich
 - stadium, *see* stadium billiard
- Burke-Shaw system
 - symmetry, 191
- Burnett coefficient, 493
- Burnett coefficients, 26, 481, 486
- butterfly effect, 32
- $C_{3vv} = D_3$ symmetry, 511
- $C_{3vv} = D_3$ symmetry, 317
- canonical transformation, 151, 152, 948
- Cartan
 - Killing classification, 151
- Cartan, É. , 248
- Cartwright, M.L., 201, 877
- cat map, 156, 165
 - 4-cycles, 375
 - Adler-Weiss generating partition, 287
 - periodic orbits count, 374
 - symbolic dynamics, 310
- Cauchy criterion, 1122
- Cauchy-Green strain tensor, 126
- caustic, 742
- ceiling function, 428, 572
- center, 101
- center manifold, 624
- center of mass, 145
- central limit theorem, 997
- central moment, 995
- centralizer, 185, 214
- chain rule, matrix, 1120
- chain-recurrent, 44
- change
 - of coordinates, 48
- chaology, *see* chaos
- chaos, 6–8
 - caveats, 10
 - deterministic, 30
 - diagnostics, 158
 - quantum, 31
 - skeleton of, 13, 15, 20, 21, 31, 420, 428, 840, 879
 - spatiotemporal, 51, 610
 - successes, 10
- character, 529, 539
 - orthonormality, 529
 - representation, 956
 - table, 956
 - D_2 , 1053
 - dihedral group, 519, 523
- characteristic
 - equation, 933
 - exponent, 127
 - function, 382, 735, 997, 1066
 - multiplier, 111, 130
 - polynomial, 356, 934, 958
 - state function, 406, 414, 998
 - value, 127
- chart, 41, 305
- section border, 73
- chicken heart palpitations, 7
- circle group, *see* $SO(2)$
- circle map, 277, 284, 338, 377, 492, 495, 649, 650, 655, 662, 665, 842,

- 895, 1043, 1047, 1086
- critical, 651, 657
- class, 183
- class algebra, 534
- Clebsch-Gordan
 - coefficients, 926
 - series, 528
- closed orbit, 119, *see* periodic orbit
- co-moving frame, 114
- coarse-graining, 382
- cocyle, 98
- coding, *see* symbolic dynamics
 - non-singular, 274
- collective excitations, 997
- collinear helium, 715
 - symbolic dynamics, 807
- combinatorics
 - teaching, 269
- compact
 - group, 199
 - invariant set, 42, 183
- complete
 - N -ary dynamics, 344
 - symbolic dynamics, 344
- completeness relation, 314, 925, 926, 934, 959
- complex
 - Ginzburg Landau equation, 228
- complex eigenvalues, 928, 936
- complexity
 - algorithmic, 364, 467
- computational degrees of freedom, 52, 53
- confession
 - C.N. Yang, 388
 - Kepler, 873
 - St. Augustine, 382
- configuration space, 51, 57
- conjugacy, 919
 - invariant, 116
 - smooth, 48, 116, 921
 - topological, 269
- conjugate momentum, 147
- conjugate, hermitian, 954
- connection formulas, 731
- connection, method of, 248
- conservation
 - equation, 672
 - phase-space volume, 148, 153, 154, 162, 392
- continuity equation, 391, 392, 672, 677, 740
- contour integral, 1009
- contracting
 - Floquet multipliers, 112, 421
 - flow, 44, 63, 95
 - map, 108, 288
 - state space, Rössler, 107, 474
- convergence
 - abscissa of absolute, 464
 - abscissa of abysmal, 464
 - radius, 433
 - super-exponential, 560, 695
- convexity, 414
- coordinate
 - change, 48, 919
 - longitudinal, 749
 - transformations, 65
- Copenhagen School, 846, 891
- correlation
 - decay
 - power law, 583
 - function, 569
 - spectrum, 569
 - time, 547
- correlation coefficients, 1002
 - matrix, 1002
- coset, 182
- cost function, 696
- counter-clockwise rotation, 229
- covariance matrix, 996
- covariant Lyapunov vector, 130, 131, 640
- covariant vector, 110, 130, 929
- covering
 - symbolic dynamics, 273
- creeping
 - 1-disk, 794
- critical
 - point, 122, 267, 270, *see* equilibrium point
 - value, 267, 492
- cross-section, 247, 248
- cumulant, 459, 468, 995, 997, 1005
 - expansion, 354, 356, 357, 365, 378, 448, 454, 997, 1000
 - Plemelj-Smithies, 1127
 - generating function, 995, 1005
- curvature
 - correction, 450
 - expansion, 29, 451
- cycle, 119, *see* periodic orbit
 - Rössler flow, 142
 - expansion, 20, 450, 763
 - 3-disk, 472
 - finite subshift, 461
 - Lyapunov exponent, 471
 - stability ordered, 462
 - fundamental, 356, 380, 450

- limit, 44, 112, 119, 124
- Lyapunov exponent, 127
- marginal stability, 115, 223, 532
- prime, 137, 273, 321, 376, 423
 - 3-disk, 330, 703
 - Hénon map, 707
- pruning, 372
- Rössler flow, 144, 333
- self-retracing, 809
- stability, 109–118
 - Gauss map, 659
- stable, 122
- superstable, 122
- weight, 435
- cycle point, *see* periodic point
- cyclic
 - group, 188
 - invariance, 137, 321
 - permutation matrix, 1019
 - symmetry, 368
- cyclic group
 - C_n
 - Character table, 1056
- cylinder set, 274, 385
- d , *see* shift
- damped Newton method, 139
- Danish pastry, *see* symbol plane
- Darboux basis, 150, 157
- de Broglie wavelength, 735
- Debye approximation, 800
- decay
 - rate, 438
 - rate of correlations, 578
- decomposition
 - irreducible, 960
- defining
 - representation, 954
 - vector space, 953
- deformation gradient, 98
- degenerate
 - eigenvalues, 927
- degree of freedom, 10, 57, 147, 625, 727
- degrees of freedom
 - computational, 52
- delta function, *see* Dirac delta
 - Dirac, 720
- density, 382, 672
 - evolution, 24
 - phase space, 392
- density of states
 - average, 760
 - Green's function, 720
 - quantum, 720
- derivative cocycle, 98
- derivative, lattice, 1021
- desymmetrization
 - 3-disk, 524
- desymmetrized state space, 247
- determinant
 - for flows, 434
 - Fredholm, 1130
 - graph, 365
 - Hadamard, 433
 - spectral, 24, 354, 433
 - trace relation, 354
 - trace-class operator, 1124
- deterministic dynamics, 6, 40, 390
- deviation
 - standard, 995
- diagonalizing matrix, 958
- diagram, *see* graph
- diffeomorphism, 78
- differential equations
 - almost ordinary, 64
- diffraction
 - Green's function, 829
 - Keller, 836
 - Sommerfeld, 836
- diffusion
 - anisotropic, 673
 - anomalous, 483
 - constant, 415, 486, 493, 673
 - equation, 673, 674
 - limited aggregate, 33
 - tensor, 459, 673
- digraph, *see* directed graph
- dihedral group, 188
 - $D_n = C_{nv}$, 1056
- dike map, 271, 288
- dimension
 - box counting, 1069
 - fractal, 1069
 - generalized, 11
 - information, 1069, 1070
 - intrinsic, 10, 625
 - physical, 120
 - symplectic, 156
- Dirac delta, 22, 23, 358, 384, 399, 407, 422, 429, 439, 563, 674, 721, 744, 755
 - derivatives, 399
 - Jacobian, 391
- Dirac path integral, 753
- discrete
 - Fourier transform, 1025
- dissipation
 - rate, Kuramoto-Sivashinsky, 636
 - rate, Navier-Stokes, 636

- dissipative
 - map, 108, 288
- divergence rate, local, 417
- divergence ultraviolet, 761
- DLA, 33, *see* diffusion limited aggregate
- dot product, 216
- doubling map, 277, 284, 565
- drift, along group tangent, 213
- dual
 - representation, 923, 953, 954
 - space, 923, 953
 - vector space, 953
- Duffing oscillator, 57, 62, 81, 161
- dyadic map, 277
- dyadic transformation, 277
- dynamical
 - localization, quantum, 156
 - transitivity, 336
 - zeta function, 19, 436
 - Euler product rep., 436
- dynamical system, 39, 41
 - axiom A, 465, 571, 943
 - equivalent, 916
 - gradient, 64
 - hyperbolic, 445, 566
 - infinite, 636
 - smooth, 20, 21, 29, 41, 360, 877, 1089, 1092, 1093
- dynamics, 40
 - are, 33
 - deterministic, 6, 40
 - hyperbolic, 345
 - irreversible, 45
 - is, 33
 - reversible, 45
 - spatiotemporal, 31
 - stochastic, 7
 - symbolic, 12, 55, 261, 272
 - symmetry, 181, 212, 235, 530
 - topological, 261, 272, 274, 337
- edge, 338
- eigendirection, 89
- eigenfunction
 - Perron-Frobenius operator, 556
 - energy, 718
 - Perron-Frobenius, 575
- eigenstate, *see* eigenfunction
- eigenvalue, 438
 - Perron-Frobenius operator, 556
 - complex, 928, 936
 - degenerate, 927
 - exponential spacing, 444
 - zero, 733, 747
- Einstein diffusion formula, 674, 684
- Einstein, A., 897
- elastic
 - scattering, 770
- elliptic
 - stability, 162
- empirical mean, 996, 1001, 1003, 1005
- enemy
 - thy, 583
- energy
 - Kuramoto-Sivashinsky, 636
 - Navier-Stokes, 636
- English
 - plain, 225, 273
- ensemble
 - microcanonical, 417
- entire function, 556
- entropy
 - barrier, 464, 881
 - Gauss map, 669
 - Kolmogorov, 173, 365, 878, 1064, 1066, 1071, 1073
 - topological, 7, 350, 358, 365
- equations of variations, 87
- equilibrium, 219
 - Rössler flow, 65
 - Kuramoto-Sivashinsky, 634
 - Lorenz flow, 62, 81
 - point, 46, 110, 387, 621, 698
 - Rössler flow, 63, 103, 282
 - relative, 219
- equilibrium measure, *see* natural measure
- equivalence
 - of dynamical systems, 916
- equivariant, 181, 182, *see* relative
 - state space, 238
 - two-mode flow, 232
- ergodic
 - average, 389
 - theorem
 - multiplicative, 130
 - theory, 389
- ergodicity, 644
- error correlation matrix, 127, 682
- error matrix, 98
- escape rate, 15, 400, 412, 416, 418, 437, 456, 468, 471, 472, 541, 545, 557, 579, 677, 946, 1009, 1066, 1070, 1071, 1133
 - 3-disk, 460, 473, 543
 - intermittency, 591
 - vanishing, 400, 460, 983
- escape time weighting, 465
- essential

- spectral radius, 562, 570
- spectrum, 562
- Euler
 - formula, 91, 576
 - limit, 92
 - MacLaurin formula, 578
 - product, 92, 437
 - product rep.
 - dynamical zeta function, 436
 - totient function, 652
- Eulerian coordinates, 47, 69, 88, 532
- evolution
 - group, 64
 - kernel probabilistic, 390
 - operator, 22, 407
 - quantum, 719
 - semigroup, 408
- evolution operator, 98
- expanding
 - Floquet multipliers, 112, 421
- expectation, 994
- expectation value, 405, 418
- expected value, 994
- exponent
 - Floquet, 112
- exponential
 - convergence, 433, 560
 - decay rate of correlations, 578
 - generating function, 403, 575, 995
 - of a matrix, 91
 - proliferation, 22, 364, 467, 881
- external node, 339
- extremal point, 728
- factor group, 183
- Faddeev, L. D., 1011
- false zeros, 437
- Farey
 - map, 264, 582, 585, 603
 - mediant, 653
 - series, 652
 - tree, 653
- Feigenbaum
 - δ , 429
 - period doubling, 879, 985
- Feynman diagrams
 - connected, 997
- Feynman path integral, 746, 753
- fiber, 46
- fiber bundle, 46
- Fick law, 673
- field, 41
- finite group, 188
- finite subshift
 - cycle expansion, 461
- first return time, 69, 306, 601
- fixed point, 137, 321
 - maps, 85
 - marginally stable, 582
 - subspace, 185, 214
 - under G , 185, 193, 194, 214
- Floquet
 - exponent, 99, 111, 112, 153, 622
 - matrix, 111
 - multiplier, 111, 112, 290, 421, 622, 640, 929
 - multiplier, metric invariant, 115
 - theorem, 929
 - theory, 112, 119
 - vector, 110, 130, 929
- flow, 39–55
 - invariant set, 109
 - autonomous, 46
 - contracting, 44, 63, 95
 - deterministic, 390
 - elliptic, 113
 - generator of, 391, 530, 639
 - Hamiltonian, 147, 961
 - hyperbolic, 113, 162, 438
 - incompressible, 96, 148, 392
 - infinite-dimensional, 52, 609–627
 - invariant subspace, 185
 - inverse hyperbolic, 162
 - linear, 90, 98
 - linearized, 98
 - map, 42
 - nonhyperbolic, 113
 - spectral determinant, 434
 - stability, 92
 - stationary, 46
 - stochastic, 390
 - stretch & fold, 285
 - symplectic, 113
- flow map, 57
- flows, 78
- Fokker-Planck equation, 676
- form, normal, 915
- formalism
 - thermodynamic, 1071
- Fourier
 - analysis, 441
 - mode, truncation, 52, 613
 - transform, discrete, 1025
 - transform, duality, 1040
- Fréchet derivative, 98
- fractal, 27, 30, 33, 70, 83, 143, 282, 308, 339, 381, 387, 398, 411, 465, 483, 486, 488, 570, 621, 633, 766, 883, 1007, 1047, 1067,

- 1074, 1086
- aggregates, 33
- dimension, 109, 1069
- geometry of nature, 33
- probabilistic, 33
- science, 11
- Fredholm
 - determinant, 1130
 - theory, 560
- free action, 198
- free energy, 998
- free node, *see* external node
- freezing, 247
- Frenkel-Kontorova model, 705
- frequency
 - analysis, 158
 - visitation, 388
- Fresnel integral, 691, 728, 734, 1100
- full shift, 344
- function, 78
 - L^2 square-integrable, 570
 - analytic, 569
 - group-invariant, 199
 - space, piecewise constant, 431
- functional, 389
 - composition, 45
 - Lyapunov, 44
- fundamental
 - cycle, 356, 380
 - domain, 184, 304
 - collinear helium, 807
 - matrix, 98, 926, 927, 933, 1061
- G -equivariant, 214
- G -fixed, 185, 193, 194, 214
- G -invariant
 - basis, *see* invariant polynomial basis
 - polynomial basis, 199, 244
- G -invariant
 - polynomial basis, 201, 208, 209, 248
- G_p -symmetric, 193
- Galerkin truncation, 52, 620
- Galilean invariance, 610, 635, 637
- Gatto Nero, Prof. Dr. Dr. h.c., 53, 504, 846, 885, 895
- gauge
 - fixing, 306
 - invariance, 306
- gauge fixing, 73
- Gauss map, 401, 603, 652, 653, 663
 - cycle stability, 659
 - metric entropy, 669
- Gauss shift, *see* Gauss map
- Gaussian
 - integral, 399, 482, 690, 755, 997
 - integral, d -dimensional, 690, 744
 - noise, 1104
 - probability density, 682
- general linear group, 180
- generalized
 - period, 536
 - periodic orbit, 536
- generating
 - function, 17, 352, 365, 408, 423, 492, 657, 705, 990, 1043, 1083
 - function, exponential, 403, 575, 995
 - orbit, 277
 - partition, 274
- generating function, 164
- generating partition
 - Adler-Weiss, 287
 - cat map, 287
- generator
 - anti-hermitian, 230, 231
 - Lie algebra, 151, 216
 - of flow, 391, 530, 639
- Gilmore, R., 201
- Ginzburg Landau equation, complex, 228
- $GL(d)$
 - general linear group, 180
- $GL(n, \mathbb{F})$, 952
- golden mean, 338, 377
 - pruning, 289, 345, 369, 377, 446, 972, 983, 1046
 - renormalization, 655, 663, 895
- good taste, 276
- gradient
 - algorithm, 696
 - system, 64
- Gram-Schmidt, 99
- grammar
 - symbolic dynamics, 275
- grandmother
 - of fractals, 33
- graph
 - irreducible, 338
 - root, 339
 - rooted tree, 339
 - strongly connected, 338
 - transition, 335
 - tree, 339
- Gray codes, 276
- Green's function, 90, 374, 685, 716, 721
 - analogue of, 644
 - connected, 997
 - density of states, 720
 - diffraction, 829
 - energy dependent, 720, 1112

- noisy, 1112
- quantum mechanical, 770
- regularized, 762
- scattering, 775
- semiclassical, 750, 751, 753, 756
- short distance, 750, 751
- trace, 720, 757, 765
 - long orbits, 750
- Greene's residue criterion, 162
- group, 951
 - compact, 199
 - cyclic, 188
 - dihedral, 188
 - dynamical, 45
 - evolution, 64
 - finite, 179, 188
 - general linear, 952
 - integral, 527–529
 - integration, 530
 - Lie, 151, 215
 - matrix, 180
 - not a, 965
 - orbit, 184, 200, 214, 219
 - orbit, marginal eigenvalue, 223
 - orbit, slice, 238, 240, 246
 - orbit, velocity, 930
 - order of, 179, 952
 - representation, 179, 530, 955
 - semi-, 384, 639
 - symmetric, 188
 - symmetry, reduction, 237, 247
 - tangent field, 216, 218
- group-invariant
 - function, 199
- Gutzwiller
 - path integral, 753
 - trace formula, 713, 760
- Gutzwiller, M., 880
- Haar measure, 530, 539, 541
- Hadamard determinant, 433
- Hadamard product, 762
- Hamilton
 - Cayley theorem, 925, 934
 - Jacobi equation, 683, 736, 751, 1111
 - equations, 736
 - principal function, 683, 739
- Hamiltonian, 737
 - dynamics, 146–154
 - equations, 147
 - flow, 147, 961
 - spectral determinant, 444
 - stability, 151, 949
 - flows, stability, 947
- Hénon map, 162
 - map
 - spectral determinant, 444
 - matrix, 151, 156
 - operator, 1119
 - repeller, periodic orbits, 144
 - separable, 718
- Hankel
 - function, 751, 784, 800
 - singular values, 125
- Hannay-Ozorio de Almeida sum rule, 550, 881
- harmonic oscillator, 921
- Harter, W. G., 534
- Harter, William G., 535
- heat
 - equation, 673, 674
- Heaviside function, 720
- Heisenberg, 892
 - picture, 1121
- Heisenberg, W., 892
- helium, 803, 891
 - collinear, 66, 84, 161, 715, 823
 - cycles, 333, 823
 - eigenenergies, 823
 - fundamental domain, 807
 - Poincaré section, 823
 - stabilities, 823
 - stability, 145
- Helmholtz equation, 784
- Helmholtz free energy, 998
- Hénon map, 78, 82, 162
 - attractor, 135, 398
 - cycles, 144, 696
 - fixed points, 84, 315
 - Hamiltonian, 162
 - horseshoe, 315
 - inverse, 315
 - Lyapunov exponent, 135
 - natural measure, 387
 - prime cycles, 707, 710
 - pruning front, 310
 - stability, 106, 122
 - structural stability, 318
 - symmetries, 961
 - time delay map, 331
 - transient, 707
- Hénon, M., 79
- Hénon-Heiles
 - symbolic dynamics, 201
- hermitian
 - conjugation, 954
 - matrix, 954
- heroes
 - unsung, xviii, 846

- Hessian matrix, 151
- heteroclinic
 - connection, 266, 274, 283, 885
 - intersection, 309
 - orbit, 105, 166, 201, 266, 277, 308
- high-dimensional
 - state space, 50
- Hilbert
 - basis, *see* invariant polynomial basis
 - space, 718
- Hilbert-Schmidt
 - operators, 1123
 - condition, 560
- Hilbert-Weyl theorem, 199
- Holmes, P., 627
- homoclinic
 - connection, 274
 - orbit, 62, 266
 - point, 874
 - tangency, 291, 302
 - tangle, 309, 661
- Hopf bifurcation, 649
- Hopf's last hope, 887
- Hopf, Eberhardt, 649, 885, 887
- Hopf, Heinz, 820
- horseshoe, 295
 - complete, 297
- Hossenfelder, S., 32
- hydrodynamic modes, 131, 1041
- hydrodynamical
 - interpretation of QM, 752
- hyperbolic
 - flow, 113, 162, 438
 - non-, 25
 - orbit, partially, 113
 - systems, 404, 445, 456, 460, 476, 483, 545, 566, 572, 592, 714, 766, 883, 896, 1067, 1071, 1115
- hyperbolicity assumption, 17, 422
- hyperchaos, 9
- iid, 996
- image space, 245, 247
- in
 - node, 102, 110
 - spiral, 102, 110
- in-degree, 339
- in-slice
 - velocity, 240
- in/out nodes, 101
- inadmissible symbol sequence, 274
- incommensurate, 43
- incompressible flow, 96, 148
- indecomposability, 336
 - metric, 263
- independently identically distributed, 996
- index
 - Maslov, *see* topological index
 - subgroup, *see* multiplicity
- index summation, repeated, 95, 152, 952
- index theory
 - Atiyah-Singer, 1011
- indicator, 382, 385
- indifferent stability, 89
- induced map, 595
- inertial manifold, 96, 620, 627
- infinite-dimensional state space, 50
- infinite-dimensional flows, 52, 609–627
- inflection point, 650
- information dimension, 1069, 1070
- initial
 - conditions, sensitivity to, 7
 - point x_0 , 16, 42, 87
 - state x_0 , 16, 42
- injective, 83
- integrable system, 148, 909
- integration
 - by parts, lattice, 1024
 - group, 530
 - Runge-Kutta, 64
- integrator, 65, 68, 84, 145, 620
 - Runge-Kutta, 1062
 - symplectic, 55, 1060–1062
- intermittency, 173, 401, 462, 463, 469, 483, 489, 565, 581, 714, 817, 1071, 1080
 - anomalous diffusion, 483
 - escape rate, 591
 - piecewise linear model, 584
 - resummation, 596
 - stability ordering, 599
- internal node, 339
- invariance
 - cyclic, 137, 321
 - Galilean, 610, 635
 - local Galilean, 637
 - of flows, 114
 - symplectic, 147, 155, 947
- invariant, 954
 - density, *see* natural measure
 - matrix, 954
 - measure, 387
 - measure, Gauss map, 401
 - metric, 111, 116
 - points, 185, 214, 504
 - polynomial basis, 199, 201, 208, 209, 244–248
 - set, compact, 42, 183

- subgroup, 183
 - subspace, 185
 - topological, 111
 - tori, 224
 - vector, 954
- inverse
 - hyperbolic, 112, 162
 - iteration, 137
 - iteration, Hamiltonian repeller, 144
- inversion, 188
- involution, 188
- inward/outward spirals, 101
- irreducible
 - decomposition, 960
 - graph, 338
 - matrix, 336
 - representation, 503, 528
 - segment, 196
- irrep, 503, 960
 - cyclic group, 518
- irrep basis
 - dihedral group, 519
- irreversibility, 24, 45
- Ising model, 316, 510, 524, 1078, 1080, 1090, 1091, 1097
- isotropy
 - subgroup, 191, 193, 194, 200, 209, 225
- isotypic decomposition, 186
- iteration, 41
 - inverse, 137
 - Hamiltonian repeller, 144
 - map, 76
- itinerary, 12, 14, 262, 317, 330
 - bi-infinite, 264, 273
 - future, 268, 273
 - past, 273
- Jacobi, C.G.J., 98
- Jacobian
 - matrix, 16, 88, 98, 99, 927
- jacobian, 95, 383
 - Hénon map, 106
- jerk, 917
- Jonquière function, 486, 489, 495, 585, 607
- Jordan normal form, 934
- KAM, 156
 - tori, 581
- Kamiltonian, 152
- Kaplan-Yorke criterion, 120
- Karhunen-Loève, 294
- Karhunen-Loève transform, 647, 996
- Keller diffraction, 836
- Keller, J.B., 879
- Kepler potential, anisotropic, 201, 467, 1055
- Kepler, Johannes, 548
- Keplerian orbit, 713
- kernel
 - resolving, 561
- kicked rotor, 149, 156
- kneading
 - determinant, 277
 - sequence, 270, 288
 - theory, 270
 - value, 270, 288
- Kolmogorov entropy, 173, 365, 878, 1064, 1066, 1071, 1073
- Kolmogorov, A. N., 878
- Koopman
 - modes, 638–644
 - operator, 638, 644, 1058, 1061
- Kraichnan, Robert H., 887
- Kramers, 892
- Krein-Friedel-Lloyd formula, 777
- Kronecker delta, 923, 953
- Krylov subspace, 99
- KS, *see* Kustaanheimo-Stiefel
- Kuramoto, Y., 627
- Kuramoto-Sivashinsky
 - “derivative” form, 627
 - “integral” form, 627
 - equation, 9, 222, 249, 610, 611, 624, 627, 631–633
 - equilibria, 633, 634
 - symmetries, 610
- kurtosis, 418, 493, 996
- Kustaanheimo-Stiefel transformation, 805, 820, 823, 913
- L^2 function space, 570
- Lagrangian, 738
 - coordinates, 47, 69, 88, 532
 - frame, 114
 - manifold, 740
- laminar states, 581
- Langevin equation, 675, 684, 685
- Laplace
 - transform, 23, 358, 410, 424, 425, 432, 720, 751, 1059
 - transform, discrete, 354, 423, 606
- Laplace, Pierre-Simon de, 6
- Laplacian
 - diagonalization, 1042
 - diagonalized, lattice, 1030
 - inverse, lattice, 1024
 - lattice, 1022
 - non-local, 1042

- large deviations, 997
- last hope, Hopf's, 887
- lattice
 - configuration, 1021
 - derivative, 1021
 - derivative, backward, 1021
 - derivative, forward, 1021
 - Fourier transform, 1025
 - integration by parts, 1024
 - Laplacian, 1022
 - Laplacian, diagonalized, 1030
 - Laplacian, inverse, 1024
 - state, 1021
 - time reversal, 201
- leaf, *see* external node
- least action principle, 330, 701
- Legendre transform, 739
- Leibniz, Gottfried Wilhelm, 6
- Letellier, C., 201
- level set, 147
- Liapunov, *see* Lyapunov
- libration orbit, 809, *see* self-retracing
- Lie
 - algebra, 151, 215, 216, 924
 - bracket, 218
 - derivative, 218
 - group, 151, 215
 - product, 924
- lifetime, 15
 - matrix, 780
- limit cycle, 44, 112, 119, 124
- linear
 - algebra, 98, 99, 101, 124, 499, 534, 922, 923
 - cocyle, 98
 - flow, 90, 98
 - space, 923
 - stability, 86, 109, 621
- linearized
 - flow, 98
- link, 338, *see* edge
- Liouville
 - equation, 393
 - operator, 393
 - theorem, 148, 153, 154, 162, 392
- Liouville's formula, 95
- little group, 193, 209
- Littlewood, J.E., 877
- local
 - divergence rate, 417
 - stability, 86, 109, 621
- localization, quantum dynamical, 156
- logistic map, *see* unimodal
- longitudinal
 - coordinate, 749
- loop
 - intersecting, 356
- Lorentz gas, 469, 476, 487, 488, 497, 581
- Lorentzian, 721
- Lorenz flow, 62, 81, 104, 105, 206, 207, 210, 282
 - complex, *see* complex Lorenz flow
 - polar coordinates, 209, 966
 - proto-Lorenz, 210
 - symmetry, 189, 191
- Lorenz, E.N., 79, 201
- Louis, Joe, American boxer, 8
- loxodromic, 949
 - quartet, 153, 157
- Lozi map, 79, 83
- Lyapunov
 - characteristic numbers, 130
 - covariant vector, 130, 131, 640
 - equation, 130, 692
 - exponent, 7, 109, 124–128, 917
 - cycle, 127
 - cycle expansion, 471
 - equilibrium, 622
 - natural measure, 417
 - numerical, 128, 131
 - function, 130
 - functional, 44
 - mode, 128, 130
 - orbit, 130
 - time, 7, 10, 25, 45, 69, 394, 403, 466, 486, 617
 - vector, 130
- Lyapunov, A.M., 877
- Möbius inversion, 362, 366, 992
- Madrid, Real, 53
- Mandelbrot, Benoit B., xviii, 33, 548
- manifold, 41
 - inertial, 96
 - unstable, 291
- Manning's multiples, 975
- map, 41, 75–78
 - area preserving, 961
 - contracting, 108, 288
 - dike, 271, 288
 - dissipative, 108, 288
 - expanding, 263
 - fixed point, 85
 - Hénon, 82, 696, 961
 - Hamiltonian, 162
 - prime cycles, 707
 - Hamiltonian
 - Hénon, 162

- iteration, 76
- logistic, *see* unimodal
- Lozi, 79, 83
- once-folding, 295
- order preserving, 269
- orientation preserving, 961
- orientation reversing, 961
- quadratic, 83
- return, 17, 69, 70, 142, 265, 282, 286, 293, 306, 307, 315, 317
- sawtooth, 188, 204, 520
- stability, 93
- tent, 285
- unimodal, 267
- mapping, 78
- marginal
 - stability, 17, 89, 102, 110, 112, 223, 421, 531, 565, 581
 - cycle, 115, 223
 - fixed point, 582
- marginal stability
 - cycle, 532
- Markov
 - chain, 273
 - graph, *see* transition graph
 - matrix, 336, 398, 416, 461
 - partition, 482, 878
 - finite, 263, 345
 - infinite, 342
 - not unique, 291
- Maslov index, *see* topological index
- material invariant, 672
- Mather, *see* Aubry-Mather theory
- matrix
 - covariance, 996
 - diagonalizing, 958
 - exponential, 91, 1059
 - group, 180
 - hermitian, 954
 - invariant, 954
 - irreducible, 336
 - negative definite, 125
 - of variations, *see* stability matrix
 - positive definite, 125, 690
 - product, 924
 - rep
 - cyclic group, 517
 - representation, 180, 924
 - stability, 87, 99, 683
- Maupertuis, P.L.M. de, 157, 330, 701
- mean, 994
- measure, 382
 - continuous, 917
 - cylinder, 274, 385
 - equipartition, 274, 385
 - invariant, 387
 - natural, 79, 388, 395, 405, 475, 547, 550, 885, 896
- mechanics
 - quantum, 718
 - statistical, 24
- memory
 - m -step, 262
 - finite, 345
- method of connections, 248
- metric
 - entropy, 109
 - entropy, Gauss map, 669
 - indecomposability, 263
 - invariant, 111, 116
 - Floquet multiplier, 115
- microcanonical ensemble, 417
- Mira, C., 79
- Misiurewicz, M., 79
- mixing, 7, 8, 17, 81, 158, 389, 405, 406, 547, 578, 883
- mode, normal, 965
- modulated traveling wave, 225
- moment, 995
 - generating function, 995
 - central, 995
 - standardized, 995
- moment-generating function, 592
- monodromy matrix, 93, 117, 421, 948
- Morse index, *see* topological index
- mother
 - of fractals, 33
- moving frame, 236, 247
 - SO(2), 256
- multi-scattering matrix, 786
- multifractal, 33, 1086
- multiplicative cocycle, 98
- multiplicative ergodic theorem, 130
- multiplicative noise, 676, 679, 682, 684
- multiplicity, 182, 193, 195, 196, 201, 206, 207, 499, 507, 510
- multiplier, Floquet, 112, 290
- multiplier shooting method, 325
- multivariate normal distribution, 996
- N -disk, transition matrix, 344
- natural density, *see* natural measure
- natural invariant, *see* natural measure
- natural measure, 79, 328, 388, 395, 405, 417, 475, 547, 550, 568, 885, 896
- nature, geometry of, 33
- Navier-Stokes equation, 609
- Navier-Stokes flow

- stability, 120
- negative definite matrix, 125
- neighborhood, 86, 117
- Nero, G., *see* Gatto Nero, 885
- neutral, *see* marginal
- New York subway map, 291
- Newton method, 138
 - convergence, 139
 - damped, 139
 - flows, 139
 - optimal section, 979
- Newtonian dynamics, 146
- node, 338
 - external, 339
 - in-degree, 339
 - internal, 339
 - out-degree, 339
- noise
 - Brownian, 675
 - Gaussian, 682, 684, 1104
 - multiplicative, 676, 679, 682, 684
 - white, 682
- non-singular coding, 274
- non-wandering set, 43, 297
- nonchaotic solution, 57
- nonequilibrium, 475
- nonhyperbolic
 - flow, 113, 114
 - systems, 550, 602, 608
- norm, 1122
- normal
 - distribution, 996
 - distribution, multivariate, 996
 - divisor, 183
 - form, 915
 - Birkhoff, 1107, 1116
 - mode, 965
 - washing machine, 179
- obscure
 - foundations, 891
 - jargon, 225, 261
 - topology, 623, 632
- observable, 388, 394, 403, 421, 475, 483, 547, 592, 601, 638, 879, 1058, 1075, 1080, 1097, 1098
 - simultaneous, 959
 - vector, 418
- ODEs, 9, 50
 - almost, 64
- 1-disk
 - creeping, 794
 - scattering, 785
 - semiclassical scattering, 791
- Onsager-Machlup, 685
- open system, 15, 411, 412, 415, 438, 546, 714, 719, 757, 783, 815, 831
- operator
 - camera, 53
 - evolution, 407
 - Hilbert-Schmidt, 1123
 - just a matrix, 1022
 - Koopman, 638, 644, 1058, 1061
 - Liouville, 393
 - norm, 1121
 - Perron-Frobenius, 384, 414
 - positive, 1123
 - regularization, 1129
 - resolvent, 354, 410, 1059
 - semigroup
 - bounded, 410, 1059
 - shift, 271, 273, 1016, 1018
 - stepping, 1016
 - trace-class, 1122
- orbit, 42, 76, 183
 - closed, 119
 - group, 219
 - inadmissible, 270
 - Keplerian, 713
 - periodic, 43, 273, 450, 758, 759
 - pseudoperiodic, 536
 - relative, 222
 - returning, 757
 - space, 247
- order preserving map, 269
- ordering
 - spatial, 268, 298
- ordinary differential equations, *see* ODEs
- orientation
 - preserving map, 961
 - reversing map, 961
- orthogonality relation, 314, 925, 926, 934, 959
- Oseledec ergodic theorem, 130
- out
 - node, 102, 110
 - spiral, 102, 110
- out-degree, 339
- Palais slice, 249
- palpitations, chicken heart, 7
- paradise
 - this side of, 543
- Parseval's identity, 1029
- partial differential equations, *see* PDEs
- partially hyperbolic
 - invariant tori, 224
 - orbit, 113
- particle

- image velocimetry, 48
- partition, 262, 274
 - state space, 382
 - function, 414, 998
 - generating, 274
 - infinite, 288, 358, 365
 - Markov, 263
- passive scalar, 672
- past topological coordinate, 299
- path integral
 - stochastic, *see* Wiener integral
- PCA, 647, 996
- PDEs, 9, 41, 50
- period
 - generalized, 536
 - relation to action, 811
 - relative, 221
- period doubling
 - bifurcation, 969
- periodic
 - orbit, 13, 43, 57, 113, 119, 220, 273, 450, 758, 759
 - condition, 136, 141, 321, 695
 - extraction, 136–141, 321–328, 695–703
 - Hamiltonian repeller, 144
 - inverse iteration, 137
 - multipoint shooting, 325
 - Newton method, 138–139
 - relative, 221, 525, 533
 - relaxation algorithm, 696
 - short, 196, 225
 - theory, 225, 349, 364, 441
 - theory, alternative, 465
 - theory, The Heresy, 467
 - point, 13, 16, 21, 22, 43, 285, 968
 - admissible, 352
 - count, 361
 - symbolic label, 273
 - unstable, 15
- periodic orbit
 - generalized, 536
 - reduced, 536
- Perron-Frobenius
 - matrix, 336, 365
 - operator, 384, 414, 575
 - theorem, 568, 573, 1089
- phase
 - relative periodic orbit, 196
- phase space, 40, 148, *see* state space
 - density, 392
 - vs. state space, 57
- physical
 - dimension, 120
- physical space, *see* configuration space
- piecewise
 - constant function, 431
 - linear map, 603
 - linear map, intermittency, 584
 - linear map, repeller, 416
- pinball, *see* 3-disk
 - simulator, 176
- PIV, *see* particle image velocimetry
- plain English, 225, 273
- plane Couette flow
 - energy, 627
 - relative solutions, 228
 - stability, 930
 - symmetries, 190, 193, 201, 228
 - unstable manifold, 293
- Plemelj-Smithies cumulants, 1127
- POD, 294, 647, 996
- Poincaré
 - return map, *see* return map
 - section
 - border, *see* chart border
 - surface of section, *see* Poincaré section
- Poincaré invariants, 154
- Poincaré return map, 69, 70, 306
 - cycle, 117
 - polynomial, 76
 - stability, 93
- Poincaré section, 14, 69–75, 307, 315, 317
 - 3-disk, 169
 - border, 85
 - Hénon trick, 78
 - hyperplane, 314
- Poincaré, H., 4, 8, 16, 57
- Poincaré-Birkhoff theorem, 164
- Poincaré section
 - hyperplane, 71
- point
 - non-wandering, 43
 - periodic, 13
 - symbolic label, 273
 - scatterer, 839
 - wandering, 43
- point-wise invariant, *see* G -fixed
- Poisson
 - bracket, 218, 392, 393, 395, 947
 - resummation, 23, 597
- polar coordinates, 921
- polar decomposition, 126, 131, 135
- Pollicott, M., 414, 601
- polylogarithm, 585
- polynomial

- characteristic, 356
- topological, 357
- Pomeau, Y., 79
- pornography, 8, 626
- positive definite matrix, 125, 690
- positive operators, 1123
- post-processing, 239, 246
- potential problems, 64
- power law
 - correlation decay, 583
- pressure
 - thermodynamic, 414
 - topological, 414
- prime cycle, 137, 273, 321, 376, 423
 - 3-disk, 315, 376, 703
 - binary, 303, 315, 369
 - count, 362
 - Hénon map, 707, 710
 - ternary, 316
- prime periodic orbit, *see* prime cycle
- primitive cycle, *see* prime cycle
- principal
 - axes, 996
 - component analysis, 647, 996
 - directions, 126
 - stretches, 126
- probabilistic zeta function, 601
- probability
 - density, Gaussian, 682
 - matrix, 336
- product
 - Lie, 924
 - matrix, 924
- professor
 - does not know this, 448
- profile, spatial, 41
- projection
 - operators, 957
- projection operator, 931, 934, 958, 1026, 1039
 - complete, orthonormal, 1026
- projection operators
 - C_n , 1023
 - C_n , 1039
- propagator, 719
 - scalar, 721
 - semiclassical, 743
 - short time, 744, 750
 - Van Vleck, 745
- proper orthogonal decomposition, 647, 996
- pruned
 - block, 275
- pruning, 12, 581
- front, 301
 - 3-disk, 264, 320
- golden mean, 289, 345, 369, 377, 446, 972, 983, 1046
- individual cycles, 372
- primary interval, 271
- rules, 344
- symbolic dynamics, 274
- pseudo-cycle, 449
- pseudoperiodic orbit, 536
- quadratic map, 83
- quantization
 - Bohr-Sommerfeld, 713
 - semiclassical, 757
 - WKB, 725, 728
- quantum
 - chaology, *see* chaos, quantum
 - chaos, 715, 716, 762
 - dynamical localization, 156
 - evolution, 719
 - interference, 735
 - mechanics, 718
 - potential, 752
 - propagator, 719
 - resonances, 713
 - theory, old, 891
- quasiperiodic, 43, 536
- quotient
 - group, 183
 - space, 247
 - state space, 73, 184
 - state space, 247
- radius of convergence, 433
- random
 - walk, 674
- random matrix theory, 716
- Rayleigh-Benard flow, 62
- real
 - Madrid, 53
- recoding, 275, 303, 315
- rectangle, 298
- rectification
 - flows, 909
 - maps, 914
- recurrence, 43, 261
 - time, *see* return time
- reduced
 - periodic orbit, 536
- reduced state space, 184, 237, 238, 247
- reflection, 188
- regular
 - group action, 238
 - representation, 502

- regular rep
 - cyclic group, 517
 - dihedral group, 518
- regularization, 762, 912
 - Green's function, 762
 - operator, 1129
- relative
 - equilibrium, 219
 - orbit, 222
 - period, 221
 - periodic orbit, 221, 525, 533
 - phase, 196
 - shift, 196
 - solutions, 425
- relaxation algorithm, 696
- renormalization, 173
 - golden mean, 655, 663, 895
- repeated index summation, 95, 152, 952
- repeller, 15, 44, 411, 715
 - piecewise-linear, 416
 - single fixed point, 575
- representation, 180
 - character, 529, 956
 - defining, 954
 - dual, 923, 953, 954
 - equivalent, 955
 - faithful, 955
 - irreducible, 503, 528
 - linear, 214
 - matrix, 530, 924, 955
 - regular, 502
 - space, 953
 - standard, 953
 - trivial, 528
- representative point, 40
- residue, 164
 - Greene's, 162
 - stability, 156, 162
- resolvent
 - kernel, 561
 - operator, 354, 410, 1059
- resonances
 - complex, 714
 - quantum, 713
 - Ruelle-Pollicott, 414, 601
- resummation
 - intermittency, 596
- return map, 17, 142, 307, 315, 317
 - Rössler flow, 282
- return time, 601
 - distribution, 601
- returning orbit, 757
- reversible
 - dynamics, 45
- Riemann zeta function, 495, 603
- right-handed rotation, 229
- Rolling Stones, 270
- root, 339
- rooted tree graph, 339
- Rössler
 - attractor, 71, 81
 - cycles, 142, 144, 333
 - equilibria, 65, 103, 282
 - flow, 57, 63, 65, 71, 81, 84, 85, 107, 128
 - Lyapunov exponent, 135
 - return map, 282
 - web diagram, 286
- rotating wave, 219, 228
- rotation
 - counter-clockwise, 229
 - right-handed, 229
- Roux, Henriette, 30, 99
- Ruelle
 - Pollicott resonances, 414, 601
 - zeta function, *see* dynamical zeta function
- Ruelle, D., 31, 414, 441, 601, 1071
- Runge-Kutta
 - integrator, 1062
- Runge-Kutta integration, 64
- running orbit
 - Lorentz gas, 479
- Rutherford, 803
- Rydberg series, 813
- S^1 , *see* $SO(2)$
- saddle, 101, 102, 110, 112
- saddle node bifurcation, 969, 970
- saddle point, *see* stationary phase
- saddle-node bifurcation, 79
- sample variance, 997
- satisfaction, 1000
- sausage, $(N+1)$ -dimensional, 213
- sawtooth map, 188, 204, 277, 520
- scalar
 - propagator, 721
- scalar multiplication, 923
- scattering
 - 3-dimensional spheres, 173
 - elastic, 770
 - Green's function, 775
 - matrix, 771
 - phase shift, 778
 - point, 839
- schmactals, *see* fractal
- Schrödinger
 - equation, 718
 - equation, time-independent, 718

- picture, 1121
- Schur's Lemma, 528
- Schwartzian derivative, 917
- section
 - optimal, 979
 - stroboscopic, 68
- section, Poincaré, 14, 69, 169, 306, 307, 317
- secular equation, 933, 934, 958
- Selberg zeta, 419, 441, 514, 714, 766, 879, 1115
- self-retracing cycle, 809
- self-similar, 22
 - fractal, 339
- SEM, 1004
- semiclassical
 - approximation, 736, 1100
 - field theory, 1100
 - Green's function, 751, 756
 - propagator, 743
 - quantization, 757
 - resonances
 - 3-disk, 1117
 - spectral determinant
 - collinear helium, 815
 - wave function, 741
 - semiclassical zeta function, 763
- semigroup, 384, 639
 - dynamical, 45, 64, 92, 93, 111, 114, 130, 170
 - evolution, 408
 - operator, 410, 1059
- sensitivity to initial conditions, 7, 32, 38, 124
- set, non-wandering, 297
- shadowing, 20, 360, 471
 - 3-disk, 460
- shift, 273
 - Bernoulli, 277, 284, 572, 575, 606, 971
 - finite type, 275
 - full, 273, 344
 - map, 650
 - operator, 271, 273, 1016, 1018
 - relative periodic orbit, 196
 - sub-, 274
- short periodic orbit, 196, 225
- similarity transformation, 49, 116
- simultaneous observables, 959
- Sinai, Ya. G., 31, 878
- Sinai-Bowen-Ruelle measure, *see* natural measure
- single fixed point
 - repeller, 575
- spectral determinant, 576
- singlet, 528, 538
- singular
 - value, 996
 - value decomposition, 131, 647, 996
 - values, 125, 130, 134
- singularity
 - branch cut, 589
- sink, 44, 102, 110, 112
- Sivashinsky, G.I., 627
- skeleton of chaos, 13, 15, 20, 21, 31, 420, 428, 840, 879
- skewness, 995
- slice, 214, 238, 247, 248
 - condition, 238
 - linear, 238
 - Palais, 249
- Smale
 - horseshoe, 310
- Smale, S., 11, 31, 290, 310, 365, 441, 877
 - Newton's method, 138
 - wild idea, 434, 440
- small divisor problem, 430
- S -matrix, 771
- smooth, 215
 - conjugacy, 48, 115, 116, 921
 - dynamical system, 636
 - dynamics, 20, 21, 29, 41, 360, 877, 1089, 1092, 1093
 - dynamics, spectral determinant, 572
 - interaction, 1095
 - map, 969
 - potential, 173
- $SO(2)$, 215, 229, 239, 256
- $SO(2)$, 102, 228, 229, 231–233, 258, 921
 - irreducible representation, 229
- $SO(3)$, 215, 890
- solution
 - symmetry, 218
- Sommerfeld
 - diffraction, 836
- source, 102, 110, 112
- $Sp(d)$
 - symplectic group, 151
- space
 - analytic functions, 569
 - average, 388, 404, 617
 - configuration, 51, 57
 - defining vector, 953
 - density functions, 431
 - dual, 923, 953
 - linear, 923
 - phase, 57

- state, 57
- vector, 923
- spaghetti, 56
- span, 923
- spatial profile, 41
- spatiotemporal chaos, 51, 610
- spatiotemporal dynamics, 31
- spectral
 - decomposition, 926, 935, 951, 959
 - determinant, 24, 354, 433
 - 1-dimensional maps, 444
 - 1D hyperbolic Hamiltonian map, 444
 - 2D hyperbolic Hamiltonian flow, 444
 - 1-degree of freedom, 764
 - 2-degrees of freedom, 765
 - entire, 444, 559
 - for flows, 434
 - infinite product rep., 435
 - single fixed point, 576
 - weighted, 445
- radius, 562, 576
 - essential, 570
- stability, 483
- staircase, 720
- spectrum
 - Balmer, 713
- specular reflection, 168
- speed, 46
- Spiegel, Edward A., 887
- SRB measure, *see* natural measure
- St. Augustine, 382
- stability, 86–96
 - billiards, 122, 170
 - continuous symmetry, 223
 - eigenvalue, *see* Floquet multiplier
 - elliptic, 430
 - equations, 98
 - exact, 123
 - exponent, 88, 95, 97, 102, 104, 110, 130, *see* Floquet exponent
 - flow, 92
 - Hamiltonian flow, 949
 - Hamiltonian flows, 151, 947
 - in-slice, 256
 - indifferent, 89
 - linear, 86, 109, 621
 - maps, 93
 - marginal, 89, 223, 531
 - matrix, 87, 98, 99, 683
 - matrix, symplectic, 151
 - multiplier, 88, 97, 130, *see* Floquet multiplier
 - neutral, *see* marginal
 - ordering
 - cycle expansions, 462
 - intermittent flows, 599
 - Poincaré map cycle, 117
 - Poincaré return map, 93
 - residue, 156, 162
 - spectral, 483
 - structural, 297, 308, 361, 483
 - subgroup, 193, 209
 - window, 113
- stabilizer, 193, 194, 225
- stabilizer subgroup, *see* isotropy subgroup
- stable
 - cycle, 122
 - manifold, 17, 291–293, 315
- stable manifold, 314, 318
- stadium billiard, 168, 173, 176, 201, 607, 1053, 1055, 1073
- stagnation point, *see* equilibrium point
- staircase
 - mean eigenvalue density, 815
 - spectral, 720
- standard
 - basis, 923
 - deviation, 995
 - map, 68, 149, 156, 163, 164, 479, 581, 843
 - representation space, 953
- standard error of the mean, 1004
- standardized moment, 995
- standing orbit
 - Lorentz gas, 479
- standing wave, 46, 228
- state, 41, 261, 338
 - set, 262
- state space, 40, 41, 148
 - discretization, 52, 414
 - equivariant, 238
 - Fourier representation, 52
 - high-dimensional, 50
 - infinite-dimensional, 52
 - partition, 382
 - reduced, 184, 238
 - visualization, 51
 - volume $\mathcal{M}$, 411
 - vs. phase space, 57
- stationary
 - flow, 46
 - phase, 390, 703, 728, 731, 733, 747, 755, 783, 797, 828, 830, 1099, 1110
 - phase approximation, 728, 734, 745, 758, 829, 839

- point, *see* equilibrium point
- state, 387
- stationary Lyapunov basis, 130
- statistical mechanics, 24
- steady state, *see* equilibrium point
- stepping operator, 1016
- Sterling formula, 734
- stochastic
 - dynamics, 7, 390
 - matrix, 336
 - path integral, *see* Wiener integral
- Stokes theorem, 154, 740
- stosszahlansatz, 25, 486
- strange attractor, 31, 44, 63, 124
 - Rössler flow, 71, 81
- stratum, 194
- stress, 1000
- stretch & fold, 82, 285
- stretches, 126
- stroboscopic sections method, 68
- strongly connected graph, 338
- structural stability, 27, 297, 308, 361, 483, 487, 614, 842, 894
 - Hénon map, 318
- structure constant, 924
- subgroup
 - isotropy, 191, 193, 209
- subshift, 274
 - finite type, 275, 301, 308, 337–339, 344
- sum
 - Birkhoff, 403, 404, 408, 414, 417, 422, 437, 450
- super-exponential
 - convergence, 695
- superstable
 - cycle, 122
 - fixed point, 695
 - point, 970
- surjective, 83
- survival probability, 15, *see* escape rate
- SVD, 647, *see* singular value decomposition, 996
- symbol
 - sequence
 - inadmissible, 274
 - square, 298
- symbol square, 298
- symbolic dynamics, 12, 55, 261–275, 968–975, 981–988
 - 3-disk, 38, 263, 319
 - at a bifurcation, 173
 - binary
 - collinear helium, 807
 - coding, 275
 - transition graph, 461
 - complete, 276, 285, 297, 344
 - covering, 273
 - grammar, 275
 - Hénon-Heiles, 201
 - pruned, 274
 - recoding, 275, 303, 315
 - unimodal, 268
- symmetric group, 188
- symmetry
 - D_3 , 317, 511
 - 3-disk, 189, 190, 205–207, 304, 317, 511, 524
 - continuous, 212–223, 235–245
 - cyclic, 368
 - discrete, 179–186, 192–199, 303, 315
 - dynamical system, 181, 212, 235, 530
 - Hénon map, 961
 - of a solution, 193, 194, 200, 225
 - solution, 218
 - under G_p , 193
- symmetry-reduced space, 237, 247
- symplectic, 113
 - 2-form, 150
 - dimension, 156
 - group $\mathrm{Sp}(2D)$, 948
 - group $\mathrm{Sp}(d)$, 151
 - Hénon map, 162
 - integrator, 55, 1060–1062
 - invariance, 147, 155, 947
 - map, 152
 - transformation, 151, 152, 295, 393
- system
 - open, 411
- syzygy, 208, 209, 245
- tangent
 - bundle, 46, 87
 - field, 216
 - field, group, 218
 - linear equations, 90
 - map, *see* stability matrix
 - space, 87, 216
- tangent linear propagator, 98
- tangent map, 98
- Tauberian theorem, 603
- teaching
 - combinatorics, 269
- template, 71, 143, 282
- tent map, 285, 917, 921
- ternary
 - prime cycles, 316

- The Heresy, 467
- thermodynamic
 - formalism, 1071
 - pressure, 414
- Thom's toral automorphism, *see* cat map
- Thom-Anosov diffeomorphism, *see* cat map
- 3-body problem, 715, 803, 873, 892, 910
- 3-dimensional sphere
 - scattering, 173
- 3-disk
 - boundary orbits, 198, 509
 - convergence, 564
 - cycle
 - analytically, 332
 - count, 186, 200, 473, 513
 - expansion, 472
 - escape rate, 418, 460, 473, 543
 - fractal dimension, 1067
 - geometry, 169
 - hyperbolicity, 422
 - pinball, 5, 172, 173, 175
 - point scatterer, 839
 - prime cycles, 18, 315, 330, 376, 703
 - pruning front, 264, 320
 - semiclassical resonances, 1117
 - shadowing, 460
 - simulator, 176
 - state space, 15, 317, 1067
 - symbolic dynamics, 13, 38, 263, 319
 - symmetry, 189, 190, 205–207, 304, 317, 511, 524
- time
 - arrow of, 24
 - as parametrization, 49
 - average, 389, 414
 - ceiling function, *see* ceiling function
 - delay, Wigner, 778
 - ordered integration, 92, 97
 - turnover, 103, 104
- time reversal, 201, 206
 - lattice, 201
- time- t forward map, 57, 78, 218
- topological
 - conjugacy, 269
 - dynamics, 261, 272, 274, 275, 337
 - entropy, 7, 350, 358
 - future coordinate, 269
 - index, 742
 - topological index, 760, 892
 - invariant, 111
 - Markov chain, 273
 - parameter, 271, 970
 - polynomial, 357
 - trace formula, 354
 - transitivity, 336
 - zeta function, 357, 358
- torus, 43
- totient function, 652
- t_p cycle weight, 435
- trace
 - formula
 - deterministic, 23
 - flows, 424
 - Gutzwiller, 760
 - maps, 423, 576
 - symmetry reduced, 533
 - topological, 354, 358
 - weight, 453
 - local, 368
- trace-class operator, 788, 1122
 - determinant, 1124
- trajectory, 42, 91
 - discrete, 76
- transfer
 - map, 157
 - matrix, 157, 398, 416
 - operator, 430, 440
 - spectrum, 444
- transformation
 - canonical, 295
 - coordinate, 65
 - symplectic, 295
- transient, 43, 263, 347
- transition
 - graph, 335–342
 - infinite, 356
 - matrix, 335, 350, 365, 368
 - matrix, N -disk, 344
- transitive, 44
- transplacement gradient, 98
- transversality
 - border of, 85, *see* chart border
 - condition, 70
 - Thom's, 894
- transverse stability, 750
- traveling wave, 219, 228, *see* relative equilibrium
- tree graph, 339
- trivial
 - representation, 528
- Trotter product formula, 1121
- truncation
 - Fourier, 52, 613
 - Galerkin, 52, 620
- turbulence, 8, 11, 56, 625
 - problem of, 609

- turnback point, 294
- turnover time, 103, 104
- two-mode flow, 212, 225, 230, 231, 233, 258, 259
 - equivariant, 232
- $U(1)$, 229, 256
- Ulam map, 285, 332, 917, 921
 - skew, 398, 554
 - tent, 400, 447
- ultraviolet divergence, 761
- unimodal
 - kneading value, 288
 - map, 267
 - map, symbolic dynamics, 268
 - well ordered symbols, 288
- unimodal map
 - bifurcation sequences, 969
- unit sphere
 - volume, 66
- unstable
 - cycle, 122
 - manifold, 17, 291–293, 315
 - periodic orbit), *see* periodic orbit
 - periodic point, 15
- unstable manifold, 314, 318
- unsung
 - heroes, xviii, 846
- UPO (Unstable Periodic Orbit), *see* periodic orbit
- Ursell function, 997
- van Kampen, N. G., 685
- Van Vleck
 - propagator, 745
- variance, 995
- variational principle, 684
- vector
 - basis, 923
 - field, 45
 - field, singularities, 909
 - invariant, 954
 - observable, 418
 - space, 923
 - defining, 953
 - dual, 953
- velocity, 45
 - in-slice, 240
- velocity gradients matrix, 87, 98, 99
- vertex, 338, *see* node
- visitation frequency, 388
- visitation sequence, *see* itinerary
- volume preservation, 172
- von Neumann
 - ergodicity, 644
- Waleffe, F., 889
- walk, *see* itinerary
- wandering point, 43
- wave function
 - semiclassical, 741
 - WKB, 742
- wave, standing, 46, 228
- web diagram
 - Rössler flow, 286
- weight
 - multiplicative, 29
- well ordered symbols
 - unimodal, 288
- Wentzel-Kramers-Brillouin, *see* WKB
- Weyl
 - rule, 760
- Weyl, H., 534, 957
- white noise, 682
- Wiener integral, 683
- Wigner
 - delay time, 778
- winding number, 164, 650, 651
- WKB, 725, 736, 1099
 - connection formulas, 731
 - quantization, 725, 728
 - wave function, 742
- Yang, C.N., 388
- Young, L.-S., 79
- Z-transform, 365
- zero eigenvalue, 733, 747
- zero, false, 437
- zeta function
 - Artin-Mazur, 357
 - dynamical, 19, 436
 - probabilistic, 601
 - Ruelle, *see* dynamical
 - topological, 357, 358

Part VI

Web appendices

THE CONCEIT of this treatise is to teach you all of deterministic, stochastic and quantum chaos (or turbulence) in one go. There is so much to say, but when? So many trees are obscuring the grandeur of the forest. So whenever possible, we have moved details of a particular topic into an appendix, numbered as the corresponding chapter of the main text. Were this monograph printed and bound, this part would be left on the ChaosBook.org website, to be consulted if a deeper dive into a particular tangent is desired.

1. A brief history of chaos (appendix [A1](#))
2. Smooth conjugacies (appendix [A2](#))
3. Linear algebra, Hamiltonian Jacobians (appendix [A4](#))
4. Lyapunov exponents done right; transport of vector fields (appendix [A6](#))
5. Cycles (appendix [A16](#))
6. Counting (appendix [A18](#))
7. Diffusion (appendix [A24](#))
8. Discrete symmetries (appendix [A25](#))
9. Implementing evolution (appendix [A31](#))
10. Thermodynamic formalism (appendix [A32](#))
11. Statistical mechanics (appendix [A33](#))
12. Quantum mechanics II (appendix [A39](#))
13. Infinite dimensional operators (appendix [A40](#))

Appendix A1

A brief history of chaos

Laws of attribution

1. **Arnol'd's Law:** everything that is discovered is named after someone else (including Arnol'd's law)
2. **Berry's Law:** sometimes, the sequence of antecedents seems endless. So, nothing is discovered for the first time.
3. **Whiteheads's Law:** Everything of importance has been said before by someone who did not discover it.

— Sir Michael V. Berry

Writing a history of anything is a reckless undertaking, especially a history of something that has preoccupied at one time or other any serious thinker from ancient Sumer to today's Hong Kong. A mathematician, to take an example, might see it this way: “[History of dynamical systems.](#)” Nevertheless, here comes yet another very imperfect attempt.

A1.1 Chaos is born

I'll maybe discuss more about its history when I learn more about it.

— Maciej Zworski

(R. Mainieri and P. Cvitanović)

TRYING TO PREDICT the motion of the Moon has preoccupied astronomers since antiquity. Accurate understanding of its motion was important for determining the longitude of ships while traversing open seas.

Kepler's Rudolphine tables had been a great improvement over previous tables, and Kepler was justly proud of his achievements. He wrote in the introduction to the announcement of Kepler's third law, *Harmonice Mundi* (Linz, 1619) in a style that would not fly with the contemporary *Physical Review Letters* editors:

What I prophesied two-and-twenty years ago, as soon as I discovered the five solids among the heavenly orbits—what I firmly believed long before I had seen Ptolemy's *Harmonics*—what I had promised my friends in the title of this book, which I named before I was sure of my discovery—what sixteen years ago, I urged as the thing to be sought—that for which I joined Tycho Brahé, for which I settled in Prague, for which I have devoted the best part of my life to astronomical contemplations, at length I have brought to light, and recognized its truth beyond my most sanguine expectations. It is not eighteen months since I got the first glimpse of light, three months since the dawn, very few days since the unveiled sun, most admirable to gaze upon, burst upon me. Nothing holds me; I will indulge my sacred fury; I will triumph over mankind by the honest confession that I have stolen the golden vases of the Egyptians to build up a tabernacle for my God far away from the confines of Egypt. If you forgive me, I rejoice; if you are angry, I can bear it; the die is cast, the book is written, to be read either now or in posterity, I care not which; it may well wait a century for a reader, as God has waited six thousand years for an observer.

Then came Newton. Classical mechanics has not stood still since Newton. The formalism that we use today was developed by Euler and Lagrange. By the end of the 1800's the three problems that would lead to the notion of chaotic dynamics were already known: the three-body problem, the ergodic hypothesis, and nonlinear oscillators.

A1.1.1 Three-body problem

Bernoulli used Newton's work on mechanics to derive the elliptic orbits of Kepler and set an example of how equations of motion could be solved by integrating. But the motion of the Moon is not well approximated by an ellipse with the Earth at a focus; at least the effects of the Sun have to be taken into account if one wants to reproduce the data the classical Greeks already possessed. To do that one has to consider the motion of three bodies: the Moon, the Earth, and the Sun. When the planets are replaced by point particles of arbitrary masses, the problem to be solved is known as the three-body problem. The three-body problem was also a model to another concern in astronomy. In the Newtonian model of the solar system it is possible for one of the planets to go from an elliptic orbit around the Sun to an orbit that escaped its dominion or that plunged right into it. Knowing if any of the planets would do so became the problem of the stability of the solar system. A planet would not meet this terrible end if the solar system consisted of two celestial bodies, but whether such fate could befall in the three-body case remained unclear.

After many failed attempts to solve the three-body problem, natural philosophers started to suspect that it was impossible to integrate. The usual technique

for integrating problems was to find the conserved quantities, quantities that do not change with time and allow one to relate the momenta and positions at different times. The impossible nature of integrating the three-body problem first came from the work of astronomer and mathematician Heinrich Bruns [29] in 1887. His work showed that no algebraic integrals or conservation laws beyond energy and momentum were present in the three-body problem; however, his result did not preclude the possibility of more complicated conserved quantities. This problem was settled by Henri Poincaré and Karl Sundman in two very different ways [16, 61].

In an attempt to promote the journal *Acta Mathematica*, its founder Gösta Mittag-Leffler received permission from King Oscar II of Sweden and Norway to establish a mathematical competition. Several questions were posed (although the king would have preferred only one), and the prize of 2500 kroner would go to the best submission. One of the questions was formulated by the German mathematician Karl Weierstrass:

Given a system of arbitrary mass points that attract each other according to Newton's laws, under the assumption that no two points ever collide, try to find a representation of the coordinates of each point as a series in a variable that is some known function of time and for all of whose values the series converges uniformly.

This problem, whose solution would considerably extend our understanding of the solar system, . . .

Poincaré's submission won the prize. He showed that conserved quantities that were analytic in the momenta and positions could not exist. To show that he introduced methods that were very geometrical in spirit: the importance of state space flow, the role of periodic orbits and their cross sections, as well as homoclinic points.

The interesting thing about Poincaré's work was that it did not solve the problem posed. He did not find a function that would give the coordinates as a function of time for all times. He did not show that it was impossible either, but rather that it could not be done with the Bernoulli technique of finding a conserved quantity and trying to integrate. Integration would seem unlikely from Poincaré's prize-winning memoir, but it was accomplished by the Finnish-born Swedish mathematician Sundman. Sundman showed that to integrate the three-body problem one had to confront two-body collisions. He did that by making them go away through a trick known as regularization of the collision manifold. The trick is not to expand the coordinates as a function of time t , but rather as a function of $\sqrt[3]{t}$. To solve the problem for all times he used a conformal map into a strip. This allowed Sundman to obtain a series expansion for the coordinates valid for all times, solving the problem that was proposed by Weierstrass in King Oscar II's competition.

The Sundman's series are not used today to compute the trajectories of any three-body system. That is more simply accomplished by numerical methods or

through series that, although divergent, produce better numerical results. The conformal map and the collision regularization mean that the series are effectively in the variable $1 - e^{-\sqrt[3]{t}}$. Quite rapidly this gets exponentially close to one, the radius of convergence of the series. Many terms, more terms than any one has ever wanted to compute, are needed to achieve numerical convergence. Though Sundman's work deserves better credit than it gets, it did not live up to Weierstrass's expectations, and the series solution did not "considerably extend our understanding of the solar system." The work that followed from Poincaré did.

A1.1.2 Ergodic hypothesis

The second problem that played a key role in development of chaotic dynamics was the ergodic hypothesis of Boltzmann. Maxwell and Boltzmann had combined the mechanics of Newton with notions of probability in order to create statistical mechanics, deriving thermodynamics from the equations of mechanics. In order to evaluate the heat capacity of the most basic system, Boltzmann assumed ergodicity, which implies that dynamical systems visit every part of the phase space allowed by conservation laws equally often. This hypothesis was extended to other averages used in statistical mechanics and was called the ergodic hypothesis. It was reformulated by Poincaré to say that a trajectory comes as close as desired to any phase space point.

Proving the ergodic hypothesis turned out to be very difficult. By the end of the twentieth century, it was shown to be true for only a few systems and wrong for several others. Early on, as a mathematical necessity, the proof of the hypothesis was broken down into two parts. First one would show that the mechanical system was ergodic (it would go near any point) and then one would show that it would go near each point equally often and regularly so that the computed averages made mathematical sense. Koopman took the first step in proving the ergodic hypothesis by realizing that he could reformulate the hypothesis using recently developed Hilbert space methods [116]. This important step demonstrated that it was possible to transform a finite-dimensional nonlinear problem into an infinite-dimensional linear problem. This does not make the problem easier, but it does allow one to use a different set of mathematical tools to approach the problem. Shortly after Koopman started lecturing on his method, von Neumann proved a version of the ergodic hypothesis, giving it the status of a theorem [141]. He proved that computed averages of an ergodic mechanical system are valid mathematically. Soon afterwards Birkhoff published a much stronger version of the theorem.

chapter 19

A1.1.3 Nonlinear oscillators

The third problem that was very influential in the development of the theory of chaotic dynamical systems was the work on nonlinear oscillators. The problem is to construct mechanical models that would aid our understanding of physical systems. Lord Rayleigh came to the problem through his interest in understanding

how musical instruments generate sound. In the first approximation one can construct a model of a musical instrument as a linear oscillator. But real instruments do not produce a simple tone forever as the linear oscillator does, so Lord Rayleigh modified this simple model by adding friction and more realistic models for the spring. By using negative friction, he cleverly created two basic models for musical instruments. These models have more than a pure tone and decay with time when not stroked. In his book *The Theory of Sound*, Lord Rayleigh introduced a series of methods that would prove quite general, such as the notion of a limit cycle, which is the periodic motion of a system regardless of initial conditions.

A1.2 Chaos grows up

A professor of mine went to go hear Derrida speak once. The entire talk was about cows; everyone was flummoxed but listened carefully, and took notes about ... cows. There was a short break, and when Derrida came back, he was like, "I'm told it is pronounced 'chaos'."

— [Phil Gentry](#), unconfirmed

(R. Mainieri)

The theorems of von Neumann and Birkhoff on the ergodic hypothesis were published in 1912 and 1913. This line of inquiry developed in two directions. One direction took an abstract approach and considered dynamical systems as transformations of measurable spaces into themselves. Could we classify these transformations in a meaningful way? This led Kolmogorov to the introduction of the concept of entropy for dynamical systems. With entropy as a dynamical invariant it became possible to classify a set of abstract dynamical systems known as the Bernoulli systems. The other line of inquiry that developed from the ergodic hypothesis attempted to find mechanical systems that are ergodic. An ergodic system could not have stable orbits, as these would break ergodicity. So in 1898 Hadamard [95] published a paper with a playful title of ‘... billiards ...,’ where he showed that the motion of balls on surfaces of constant negative curvature is everywhere unstable. This dynamical system was to prove very useful and it was taken up by Birkhoff. Morse in 1923 showed that it was possible to enumerate the orbits of a ball on a surface of constant negative curvature. He did this by introducing a symbolic code to each orbit and showed that the number of possible codes grew exponentially with the length of the code. With contributions by Artin, Hedlund, and H. Hopf it was eventually proven that the motion of a ball on a surface of constant negative curvature was ergodic. The importance of this result escaped most physicists, one exception being Krylov, who understood that a physical billiard was a dynamical system on a surface of negative curvature, but with the curvature concentrated along the lines of collision. Sinai, who was the first to show that a physical billiard can be ergodic, knew Krylov’s work well.

The work of Lord Rayleigh also received vigorous development. It prompted many experiments and some theoretical development by van der Pol, Duffing, and

Hayashi. They found other systems in which the nonlinear oscillator played a role and classified the possible motions of these systems. This concreteness of experiments, and the possibility of analysis was too much of a temptation for [Mary Lucy Cartwright](#) and J.E. Littlewood [30], who set out to prove that many of the structures conjectured by the experimentalists and theoretical physicists did indeed follow from the equations of motion. Birkhoff had found a ‘remarkable curve’ in a two dimensional map; it appeared to be non-differentiable and it would be nice to see if a smooth flow could generate such a curve. The work of Cartwright and Littlewood lead to the work of Levinson, which in turn provided the basis for the horseshoe construction of S. Smale.

chapter 15

In Russia, Lyapunov paralleled the methods of Poincaré and initiated the strong Russian dynamical systems school [130]. Andronov carried on with the study of nonlinear oscillators and in 1937 introduced together with Pontryagin the notion of coarse systems. They were formalizing the understanding garnered from the study of nonlinear oscillators, the understanding that many of the details on how these oscillators work do not affect the overall picture of the state space: there will still be limit cycles if one changes the dissipation or spring force function by a little bit. And changing the system a little bit has the great advantage of eliminating exceptional cases in the mathematical analysis. Coarse systems were the concept that caught Smale’s attention and enticed him to study dynamical systems.

A1.3 Chaos with us

The past is never dead. It’s not even past.

— William Faulkner, *Requiem for a Nun*

(R. Mainieri)

In the fall of 1961 Steven Smale was invited to Kiev where he met Arnol’d, Anosov, Sinai, and Novikov. He lectured there, and spent a lot of time with Anosov. He suggested a series of conjectures, most of which Anosov proved within a year. It was Anosov who showed that there are dynamical systems for which all points (as opposed to a non-wandering set) admit the hyperbolic structure, and it was in honor of this result that Smale named these systems Axiom-A. In Kiev Smale found a receptive audience that had been thinking about these problems. Smale’s result catalyzed their thoughts and initiated a chain of developments that persisted into the 1970’s.

Smale combined their results and developments into a 1967 review article on dynamical systems, entitled “Differentiable dynamical systems” [176]. There are many great ideas in this paper: the global foliation of invariant sets of the map into disjoint stable and unstable parts; the existence of a horseshoe and enumeration and ordering of all its orbits; the use of zeta functions to study dynamical systems. The paper also emphasizes global properties of dynamical systems and the topology of orbits. Smale’s account takes you from a local differential equation (in the

chapter 15

form of vector fields) to the global topological description in terms of horseshoes.

The path traversed from ergodicity to entropy is a little more confusing. The general character of entropy was understood by Weiner, who seemed to have spoken to Shannon. In 1948 Shannon published his results on information theory [168], where he discusses the entropy of shift transformations. Kolmogorov went far beyond and suggested a definition of the metric entropy of an area preserving transformation in order to classify Bernoulli shifts [114, 115]. The suggestion was taken by his student Sinai [170] and the results published in 1959. In 1960 Rokhlin connected these results to measure-theoretical notions of entropy [153]. The next step was published in 1965 by Adler and Palis [3], and also Adler, Konheim, McAndrew [2]; these papers showed that one could define the notion of topological entropy and use it as an invariant to classify continuous maps. In 1967 Anosov and Sinai applied the notion of entropy to the study of dynamical systems [5]. It was in the context of studying the entropy associated to a dynamical system that Sinai [171, 172] introduced Markov partitions in 1968.

Markov partitions allow one to relate dynamical systems and statistical mechanics; this has been a very fruitful relationship. It adds measure notions to the topological framework laid down in Smale's paper. Markov partitions divide the state space of the dynamical system into nice little boxes that map into each other. Each box is labeled by a code and the dynamics on the state space maps the codes around, inducing a symbolic dynamics. From the number of boxes needed to cover all the space, Sinai was able to define the notion of entropy of a dynamical system. In 1970 Bowen came up independently with the same ideas, although there was presumably some flow of information back and forth before these papers got published. Bowen also introduced the important concept of shadowing of chaotic orbits. We do not know whether at this point the relations with statistical mechanics were clear to everyone. They became explicit in the work of Ruelle. Ruelle understood that the topology of the orbits could be specified by a symbolic code, and that one could associate an 'energy' to each orbit. The energies could be formally combined in a 'partition function' to generate the invariant measure of the system.

After Smale, Sinai, Bowen, and Ruelle had laid the foundations of the statistical mechanics approach to chaotic systems, research turned to studying particular cases. The simplest case to consider is 1-dimensional maps. The topology of the orbits for parabola-like maps was worked out in 1973 by Metropolis, Stein, and Stein [136]. The more general 1-dimensional case was worked out in 1976 by Milnor and Thurston in a widely circulated preprint, whose extended version eventually got published in 1988 [137].

A lecture of Smale and the results of Metropolis, Stein, and Stein inspired Feigenbaum to study simple maps. This led him to the discovery of the universality in quadratic maps and the application of ideas from field-theory to dynamical systems. Feigenbaum's work was the culmination in the study of 1-dimensional systems; a complete analysis of a nontrivial transition to chaos. Feigenbaum introduced many new ideas into the field: the use of the renormalization group which

led him to introduce functional equations in the study of dynamical systems, the scaling function which completed the link between dynamical systems and statistical mechanics, and the presentation functions which describe the dynamics of scaling functions.

The work in more than one dimension progressed very slowly and is still far from complete. The first result in trying to understand the topology of the orbits in two dimensions (the equivalent of Metropolis, Stein, and Stein, or Milnor and Thurston's work) was obtained by Thurston. Around 1975 Thurston was giving lectures "On the geometry and dynamics of diffeomorphisms of surfaces." The techniques discussed in those lectures have not yet been applied to physics, but much of the classification that Thurston developed can be obtained from the notion of a 'pruning front' formulated independently by Cvitanović.

Once one develops an understanding of the topology of the orbits of a dynamical system, one needs to be able to compute its properties. Ruelle had already generalized the zeta function introduced by Artin and Mazur [6], so that it could be used to compute the average value of observables. The difficulty with Ruelle's zeta function is that it does not converge very well. Starting out from Smale's observation that a chaotic dynamical system is dense with a set of periodic orbits, Cvitanović used these orbits as a skeleton, on which to evaluate the averages of observables, and organized such calculations in terms of rapidly converging cycle expansions. This convergence is attained by using the shorter orbits used as a basis for shadowing the longer orbits.

This account is far from complete, but we hope that it gives the reader an overview of the field. It is not a fad, and it will not die anytime soon.

A1.4 Periodic orbit theory

Pure mathematics is a branch of applied mathematics.

— Joe Keller, asked to define applied mathematics

(P. Cvitanović)

The history of periodic orbit theory is rich and curious; recent advances are equally inspired by more than a century of developments in three separate subjects: 1. *deterministic chaotic dynamics*, initiated by Poincaré and put on its modern footing by Smale [176], Ruelle [161], and many others, 2. *quantum theory* initiated by Bohr, with the modern 'chaotic' formulation by Gutzwiller [92, 94], and 3. *analytic number theory* initiated by Riemann and formulated as a spectral problem by Selberg [135, 167]. Following different lines of reasoning and driven by different motivations, the three separate roads all arrive at *trace formulas*, *zeta functions* and *spectral determinants*.



The fact that these fields are all related is far from obvious, and even today the practitioners tend to cite papers only from their sub-speciality. In Gutzwiller's

words [94], “The deterministic periodic orbits are a crucial stepping stone in the understanding of quantum mechanics, in particular when the classical system is chaotic. This situation is very satisfying when one thinks of Poincaré who emphasized the importance of periodic orbits in classical mechanics, but could not have had any idea of what they could mean for quantum mechanics. The set of energy levels and the set of periodic orbits are complementary to each other since they are essentially related through a Fourier transform. Such a relation had been found earlier by the mathematicians in the study of the Laplacian operator on Riemannian surfaces with constant negative curvature. This led to Selberg’s trace formula in 1956 which has exactly the same form, but happens to be exact.” *A posteriori*, one can say that zeta functions arise in both classical and quantum mechanics because the dynamical evolution can be described by the action of linear evolution (or transfer) operators on infinite-dimensional vector spaces. The spectra of these operators are given by the zeros of appropriate determinants. One way to evaluate determinants is to expand them in terms of traces, $\log \det(\mathcal{L}) = \text{tr}(\log \mathcal{L})$. In this way the spectrum of an evolution operator becomes related to its traces, i.e. periodic orbits. A deeper way of restating this is to observe that the trace formulas perform the same role in all of the above problems; they relate the spectrum of periods (local dynamics) to the spectrum of eigenvalues (global eigenstates), and for nonlinear geometries they play a role analogous to the one that Fourier transform plays for the circle.

section 22.1

exercise 4.1

Distant history is easily sanitized and mythologized. As we approach the present, our vision is inevitably more myopic; for very different accounts covering the same recent history, see V. Baladi [15] (a mathematician’s perspective), and M. V. Berry [22] (a quantum chaologist’s perspective). We are grateful for any comments from the reader that would help make what follows fair and balanced.

M. Gutzwiller was the first to demonstrate that chaotic dynamics is built upon unstable periodic orbits in his 1960’s work on the quantization of deterministically chaotic quantum systems, where the ‘Gutzwiller trace formula’ gives the semiclassical quantum spectrum as a sum over deterministic periodic orbits [89–92]. Equally important was D. Ruelle’s 1970’s work on hyperbolic systems, where ergodic averages associated with natural invariant measures are expressed as weighted sums on the infinite set of unstable periodic orbits embedded in the underlying chaotic set [155, 156]. This idea can be traced back to the following sources: 1. the foundational 1967 review [176], where S. Smale proposed as “a wild idea in this direction” a (technically incorrect, but prescient) zeta function over periodic orbits, 2. the 1965 Artin-Mazur zeta function for counting periodic orbits [6], and 3. the 1956 Selberg number-theoretic zeta functions for Riemann surfaces of constant curvature [167]. That one could *compute* using these infinite sets was not clear at all. Ruelle [161] never attempted explicit computations, and Gutzwiller only attempted to implement summations over anisotropic Kepler periodic orbits by treating them as Ising model configurations [93] (In retrospect, Gutzwiller was lucky; it turns out that the more periodic orbits one includes, the worse convergence one gets [31]).



chapter 39

chapter 22

remark 22.2

chapter 18

For a long time the convergence of such sums bedeviled the practitioners, until the mathematically rigorous spectral determinants for hyperbolic deterministic

flows, and the closely related semiclassical exact Gutzwiller Zeta functions were recast in terms of highly convergent *cycle expansions*. Under these circumstances, a relatively few short periodic orbits lead to highly accurate long time averages of quantities measured in chaotic dynamics and of spectra for quantum systems. The idea, in a nutshell, is that long orbits are shadowed by shorter orbits, and the n th term in a cycle expansion is the difference between the shorter cycles estimate of the period n -cycles' contribution and the exact n -cycles sum. For unstable, hyperbolic flows, this difference falls off exponentially or super-exponentially [162]. Contrary to some literature, resummations of cycle expansions are not more clever than the Plemelj-Smithies recursion formula. Those who study cycle expansion often express concerns regarding the 'abscissa of absolute convergence', the 'entropy barrier', and exponential proliferation of cycles, but these concepts are not a problem scientifically; it's simply that practitioners of cycle expansion find their own theories more reassuring.

chapter A40

Cvitanović derived 'cycle expansions' in 1986-87, in an effort to prove that the mode-locking dimension for critical circle maps discovered by Jensen, Bak and Bohr [106] is universal; the same kind of periodic orbits are involved in the Hénon map, but now in renormalization 'time'. The symbolic dynamics of the Hénon attractor (the pruning front conjecture [49]) is coded by transition graphs, and topological entropy is given by roots of their determinants. This observation led to the study of convergence of spectral determinants for both discrete-time (iterated maps) and continuous-time deterministic flows (both ODEs and PDEs). Cycle expansions thus arose not from temporal dynamics, but from studies of scalings in period-doubling and cycle-map renormalizations [9, 33, 50]. This work was done in collaboration with R. Artuso (PhD 1987-1989), G. Gunaratne, and E. Aurell (PhD 1984-1989), and it was written up under the watchful eye of parrot Gaspar in Fundação de Faca, Porto Seguro, as two long *Recycling of strange sets* papers [8, 9]: *I. Cycle expansions* and *II. Applications*. The main lesson was that one should never split theory and applications into papers numbered I and II; part II, which covers many interesting results, has barely been glanced at by anyone.

chapter 23

chapter 18

chapter 28

The first published paper on these developments was written by Auerbach *et al.* [11] and entitled *Exploring chaotic motion through periodic orbits* (submitted March 1987). Here only a 'level sum' approximation (23.39),

section 27.4

$$1 = \sum_{x_j \in \text{Fix } f^n} t_j e^{\beta A(x_j, n)}, \quad t_j = \frac{e^{-n s(n)}}{\Lambda_j}, \quad (\text{A1.1})$$

to the trace formula is presented as an n th order estimate of the leading Perron-Frobenius eigenvalue $s(n)$ and applied to the Hénon attractor, as shown in eq. (4) of the above paper. (The *exact* weight of an unstable prime periodic orbit p -for level sum (21.6)- had been conjectured by Kadanoff and Tang [107] in 1984. 1 on the LHS of (A1.1) is the statement of probability conservation. In semiclassical quantization of chaos this statement that total probability is conserved even has a name: it is called 'Hannay-Ozorio de Almeida sum rule'.) Even as it was written, the heuristics of this paper were rendered obsolete by exact cycle expansions, and yet, mysteriously, this is one of the most cited periodic orbits papers.

remark 27.2

The first attempt to make cycle expansions accessible to every person was

condensed into [Phys. Rev. Letter](#) *Invariant measurement of strange sets in terms of cycles* (submitted March 1988) [43]. However, the two long papers by Artuso *et al.* [8, 9] are a better read.

Several applications of the new methodology are worth mentioning. One was the accurate calculation of the leading dozen eigenvalues of the period-doubling operator [9, 33, 151]. Another breakthrough was the cycle expansion of deterministic transport coefficients [7, 41, 48], such as diffusion constants without *any* probabilistic assumptions. The classical Boltzmann equation for the evolution of 1-particle density is based on *Stosszahlansatz*, the assumption that velocities of colliding particles are not correlated. In periodic orbit theory all correlations are included in cycle averaging formulas, such as the cycle expansion for a particle diffusing chaotically across a spatially-periodic array.

chapter 24

Physicists tend to obsess about matters weightier than iterating maps, so in 1989 Cvitanović and Eckhardt showed that cycle expansions reproduce quantum resonances of Eckhardt's 3-disk scatterer [64] to rather impressive accuracy [45]. In that same year, Gaspard and Rice published a lovely triptych of articles about the same 3-disk system (deterministic, semiclassical and quantum scattering) [79–81]. In 1992, the PhD thesis of [P. E. Rosenqvist](#) [65, 154] combined the magic of spectral determinants with their symmetry factorizations [47, 93] to take cycle expansions to ridiculous accuracy; for example, periodic orbits up to 10 bounces determine the deterministic escape rate for a 3-disk pinball to be

$$\gamma = 0.4103384077693464893384613078192 \dots$$

Try to extract this from a direct numerical simulation, or a log-log plot of level sums (A1.1)! Prior to cycle expansions, the best accuracy that Gaspard and Rice achieved by applying Markov approximations to the spectral determinant [80] was 1 significant digit, $\gamma \simeq 0.45$.

A 3-disk billiard is exceptionally nice, uniformly hyperbolic repeller. More often than not, good symbolic dynamics for a given flow is either not available, or its grammar is not finite, or the convergence of cycle expansions is affected by nonhyperbolic regions of state space. In those cases truncations such as the *stability cutoff* of Dahlqvist and Russberg [55, 56] and Dettmann and Morriss [60] might be helpful. The idea is to truncate the cycle expansion by including only the shadow combinations of pseudo-cycles $\{p_1, p_2 \dots, p_k\}$, such that $|\Lambda_{p_1} \dots \Lambda_{p_k}| \leq \Lambda_{\max}$, with the cutoff $\Lambda_{\max}$ equal to or smaller than the most unstable Λ_p in the data set.

chapter 29

section 23.7

It is pedagogically easier to motivate sums over periodic orbits by starting with discrete time dynamical systems, but most flows of physical interest are continuous in time. The weighted averages of periodic orbits for continuous time flows were introduced by Bowen, who treated them as Poincaré section suspensions weighted by the ‘time ceiling’ function, and were incorporated into dynamical zeta functions by Parry and Pollicott [148] and Ruelle [157]. For people steeped in quantum mechanics it all looked very unfamiliar, so in 1991 Cvitanović and Eckhardt reformulated spectral determinants for continuous time flows along the lines of Gutzwiller's derivation of the semiclassical trace formula [46]. As a con-

chapter 21

sequence, quantum mechanics [22, 111, 117] tend to cite this paper as the first paper on cycle expansions.

2D billiards are only toys, but quantization of helium is surely not just a game. By implementing cycle expansions in 1991, the group of Dieter Wintgen obtained a surprisingly accurate helium spectrum [68, 191] from a small set of its shortest cycles. This happened 50 years after old quantum theory had failed to do so and 20 years after Gutzwiller first introduced his quantization of chaotic systems [92].

The Copenhagen group gave many conference and seminar talks about cycle expansions. In December 1986, Cvitanović presented results on the periodic-orbit description of the topology of Lozi and Hénon attractors and the periodic-orbit computation of associated dynamical averages, at the meeting on “*Chaos and Related Nonlinear Phenomena: Where do we go from here?*” This meeting was organized by Moshe Shapiro and Itamar Procaccia and held in the kibutz Kiryat Anavim. A great meeting, and Celso Grebogi was in the audience. After the “*Where do we go from here?*” meeting, the Maryland group wrote a series of papers on unstable periodic orbits, or ‘UPOs’. In the first paper [86], *Unstable periodic orbits and the dimensions of multifractal chaotic attractor* (submitted September 1987), the focus was on fractal dimensions of chaotic attractors, as was the fashion in the late 1980’s. They prove that the natural measure ρ_0 of a mixing hyperbolic attractor is given by the limit of a sum over the unstable periodic points x_j of long period n , embedded in a chaotic attractor. Each periodic point is weighted by the inverse of the product of its periodic orbit’s expanding Floquet multipliers Λ_j , Eq. (14) in their paper:

section 15.4

remark 5.1

$$\rho_0(\mathcal{M}_S) = \lim_{n \rightarrow \infty} \sum_{x_j \in \text{Fix } f^n} \frac{1}{\Lambda_j}, \quad x_j \in \mathcal{M}_S. \quad (\text{A1.2})$$

This is an approximate level sum formula for natural measure, a special case of (A1.1), with leading Perron-Frobenius eigenvalue $s = 0$ (no escape), and $\beta = 0$ (observable = 1). The first paper does cite Auerbach *et al.* [11], in which the same approximate level sum seems to have been published for the first time. Ever since then, various cyclist teams cite exclusively their own papers and some of the mathematicians of the 1970’s.

So you have now written a paper that uses periodic orbits. What is one to cite? Work by Sinai-Bowen-Ruelle is smarter and more profound than the vast majority of ‘chaos’ publications from the 1980s on. If you are not actually computing anything using periodic orbits and are reluctant to refer to recent contributions, you can safely credit Ruelle [156, 161] for deriving the dynamical (or Ruelle) zeta function, and Gutzwiller for formulating semiclassical quantization as a Zeta function over unstable periodic orbits [92, 94]. There are no cycle expansions in these papers or in Bowen’s work (see, for example, the description in [Scholarpedia.org](https://scholarpedia.org)). If you have computed something using sums weighted by periodic-orbit weights, cite the first paper that introduced them, as well as a useful up-to-date reference, which in this case is ChaosBook.org. Do not faint because this web-book is available on (gasp!) the internet - it’s third millennium, and having a continuously updated, hyperlinked and reliable reference has its virtues.

Depending on the context, one should also cite 1) Zoldi and Greenside [192] for being the second to determine unstable periodic orbits for Kuramoto-Sivashinsky (127 of them), on a domain larger than what was studied by Christiansen *et al.* [32], 2) López *et al.* [126] for being the first to determine relative periodic orbits in a spatio-temporal PDE (complex Landau-Ginzburg), and 3) Kazantsev [110] for being the first to determine periodic orbits in a weather model, and for formulating a variational method for finding periodic orbits. We love these authors, but not for their ‘escape-time weighting’.

remark 23.1

chapter 34

chapter 34

While derivations of (A1.1) by Kadanoff and Tang 1984 and Auerbach *et al.* 1987 were heuristic, Grebogi, Ott and Yorke 1987 prove (A1.2) by taking the $n \rightarrow \infty$ limit. In actual computations it would be madness to attempt to take such a limit, as longer and longer periodic orbits become more and more unstable, exponentially growing in number, and non-computable; and the natural measure ρ_0 is everywhere singular, with support on a fractal set, with its $n \rightarrow \infty$ limit even more impossible to compute. And why would one take this limit? The whole point of cycle expansions is that it is smarter to compute averages *without* constructing the natural measure ρ_0 .

Taking a limit to obtain a proof is good mathematics, but in statistical mechanics a partition function is not a limit of anything; it is the full sum of all states. Likewise, its ergodic theory cousin, the spectral determinant is not a long-time limit; it is the exact sum over all periodic orbits. Cycle expansions were introduced in a non-rigorous manner, on purpose [43]: the exposition was meant not to frighten a novice, innocent of Borel measurable α to Ω sets. This was set right in the elegant PhD thesis of H. H. Rugh’s in 1992, *The correlation spectrum for hyperbolic analytic maps* [162], which proves that the zeros of spectral determinants are indeed the Ruelle-Pollicott resonances [150, 158, 159]. The proof is well within mathematicians’ comfort zone, so they tend to cite Rugh’s paper as the paper on ‘Fredholm determinants’, and, as always, throw in “a sense of Grothendieck” for good measure [15, 77], without citing earlier papers on cycle expansions.

chapter 28

If you intend to determine and use periodic orbits, here is the message: Heuristic ‘level sums’ are approximations to the exact trace formulas (that are derived here, in ChaosBook, and Gaspard monograph [78] with no more effort than the heuristic approximations), not smart for computations; faster convergence is obtained by utilizing the shadowing that is built into the exact cycle expansions of dynamical zeta functions and spectral determinants. Cycle expansions are *not* heuristic, in deterministic dynamics they are *exact* expansions in the unstable periodic orbits [8, 9, 43]; in quantum mechanics and stochastic mechanics they are semiclassically exact. So why would one prefer a limit of a heuristic sum such as (A1.2) to the exact spectral determinant, convergent exact periodic orbits sums, and exact periodic orbits formulas for dynamical averages of observables? It is not even wrong. Perhaps if one is very fond of baker’s maps [145], which, being piecewise linear, have no cycle expansion curvature terms, one does not appreciate the shadowing cancelations built into the spectral determinants and their cycle expansions. That might be the reason why linear thinkers stop at the level sum (A1.2).

section 27.4

A1.5 Dynamicist's vision of turbulence

Those were days when plumbers were plumbers...

—Prof. Dr. Dr. h.c. Gatto Nero, Grande Ufficiale
dell'Ordine della Stella della Notorietà Italiana

(P. Cvitanović and L. van Veen)

The key theoretical concepts that form the basis of dynamical theories of turbulence of chapter 30 are rooted in the work of Poincaré, Hopf, Smale, Ruelle, Gutzwiller and Spiegel. In his 1889 analysis of the three-body problem [149], Poincaré introduced the geometric approach to dynamical systems and methods that lie at the core of the theory developed here: qualitative topology of state space flows, Poincaré sections, key roles played by equilibria, periodic orbits, heteroclinic connections, and their stable/unstable manifolds.

In a seminal 1948 paper [100], Ebehardt Hopf visualized the function space of allowable Navier-Stokes velocity fields as an infinite-dimensional state space, parameterized by viscosity, boundary conditions and external forces, with instantaneous flow states represented by points in this state space. Laminar flows correspond to equilibrium points, globally stable for sufficiently large viscosity. As the viscosity decreases (as the Reynolds number increases), turbulent states set in, represented by chaotic state space trajectories. Hopf's observation that viscosity causes a contraction of state space volumes under the action of dynamics led to his key conjecture, which states that long-term, typically observed solutions of the Navier-Stokes equations lie on finite-dimensional manifolds embedded in the infinite-dimensional state space of allowed states. Hopf's manifold, known today as the 'inertial manifold,' is well-studied in the mathematics of spatio-temporal PDEs. Its finite dimensionality for non-vanishing 'viscosity' parameter has been rigorously established in certain settings by Foias and collaborators [76]. Hopf presciently noted that "the geometrical picture of the phase flow is, however, not the most important problem of the theory of turbulence. Of greater importance is the determination of the probability distributions associated with the phase flow". Hopf's call for understanding probability distributions associated with the phase flow has indeed proven to be a key challenge, one in which dynamical systems theory has made the greatest progress in the last half century. In particular, the Sinai-Ruelle-Bowen ergodic theory of 'natural' or SRB measures has played a critical role in understanding dissipative systems with chaotic behavior [25, 161, 173, 176].

Hopf noted "[t]he great mathematical difficulties of these important problems are well known and at present the way to a successful attack on them seems hopelessly barred. However, there is no doubt that many characteristic features of the hydrodynamical phase flow occur in a much larger class of similar problems governed by non-linear space-time systems. In order to gain insight into the nature of hydrodynamical phase flows we are, at present, forced to find and to treat simplified examples within that class." Hopf's call for geometric state space analysis of simplified models first came to fulfillment with the influential Lorenz's truncation [127] of the Rayleigh-Bénard convection state space. The Proper Orthogonal

example 2.2

Decomposition (POD) models of boundary-layer turbulence brought this type of analysis closer to physical hydrodynamics [10, 99]. Further significant progress has proved possible for systems such as the 1-spatial dimension Kuramoto-Sivashinsky flow [119, 174], which is a paradigmatic model of turbulent dynamics, as well as one of the most extensively studied spatially extended dynamical systems.

Numerical methods for the computation of equilibria and periodic orbits in dynamical systems with few computational degrees of freedom were formulated in the 1970s, e.g. by Allgower and Georg [4], and implemented in software packages soon after, e.g. by Sebius Doedel [62, 63]. However, these methods require the computation, storage and decomposition of large Jacobian matrices, of the size of the number of computational degrees of freedom. For the Lorenz model, this could be easily accomplished on a 1981 computer, but turbulent flow is a different, well-stirred, cup of tea. Even if one fixes the Reynolds number to the lowest possible value that allows for irregular motion à la Hopf, if one exploits the incompressibility condition and spatial symmetries of the flow, one still ends up with thousands of computational degrees of freedom. Matrices that large did not fit in 1980s Random Access Memory. Until a different method, aptly labelled *matrix-free*, was introduced years later, computing invariant solutions in Navier-Stokes flow was a matter of pushing the limits and searching for the latest and largest computers.

Two pioneers in this development were Andrew Cliffe and Masato Nagata. The former wrote a code called ENTWIFE [34] that had some of the functionality of Doedel's AUTO, but was geared towards the analysis of Navier-Stokes flow discretized by finite elements. One of the early results was the computation of steady, convective solutions in Taylor-Couette flow [35]. Their results filled a number of blanks in the bifurcation diagram that starts with an instability of the laminar state. Nagata tackled an even more challenging problem, namely that of travelling waves in plane Couette flow. In this geometry, the laminar flow remains stable and, when Nagata started his quest, it was not known if any interesting invariant solutions even existed. Eventually, he managed to compute a travelling wave solution at a marginal resolution. It took him more than two years to get the result published in the Journal of Fluid Mechanics, forced as he was to seek out larger computers after each round of reviewing in order to increase the resolution and, with that, the number of computational degrees of freedom [140]. Inspired by these results, and using essentially the same methods, Fabian Waleffe computed equilibria and travelling waves in plane Couette flow and Poiseuille flow [187].

We hope to have convinced the reader that a successful approach to solving the full Navier-Stokes problem is no longer “hopelessly barred.” It is now within reach given our computational and experimental capabilities. We address this challenge in a way Hopf could not divine, employing methodology developed only within the past two decades, explained in depth throughout this book.

chapter 30

Hopf, to the best of our knowledge, never suggested that turbulent flow should be analyzed in terms of ‘recurrent flows’, i.e. time-periodic solutions of the defining PDEs. The story so far goes like this: in 1960 Ed Spiegel was Robert Kraichnan's research associate. Kraichnan told him, “Flow follows a regular solution

for a while, then another one, then switches to another one; that's turbulence." It was not too clear, but Kraichnan's vision of turbulence moved Spiegel. In 1962 Spiegel and Derek Moore investigated a set of 3rd order convection equations which seemed to follow one periodic solution, then another, and continued going from periodic solution to periodic solution. Ed told Derek, "This is turbulence!" and Derek said "This is wonderful!" He gave a lecture at Caltech in 1964 and came back very angry. They pilloried him there. "Why is this turbulence?" they kept asking and he could not answer, so he expunged the word 'turbulence' from their 1966 paper [139] on periodic solutions. In 1970 Spiegel met Kraichnan and told him, "This vision of turbulence of yours has been very useful to me." Kraichnan said: "That wasn't my vision, that was Hopf's vision." What Hopf *actually* said and where he said it remains deeply obscure to this very day. There are papers that lump him together with Landau, as the 'Landau-Hopf's incorrect theory of turbulence,' a proposal to deploy incommensurate frequencies as building blocks of turbulence. This was Landau's guess and was the only one that could be implemented at the time.

The first paper to advocate a periodic orbit description of turbulent flows is thus the 1966 Spiegel and Moore paper [139, 177]. Thirty years later, in 1996 Christiansen *et al.* [32] proposed (in what is now the gold standard for exemplary ChaosBook.org/projects) that the periodic orbit theory be applied to infinite-dimensional flows, such as the Navier-Stokes, using the Kuramoto-Sivashinsky model as a laboratory for exploring the dynamics close to the onset of spatiotemporal chaos. The main conceptual advance in this initial foray was the demonstration that the high-dimensional (16-64 mode Gal rkin truncations) dynamics of this dissipative flow can be reduced to an approximately 1-dimensional return map $s \rightarrow f(s)$, by choosing the unstable manifold of the shortest periodic orbit as the intrinsic curvilinear coordinate from which to measure near recurrences. For the first time for any nonlinear PDE, some 1,000 unstable periodic orbits were determined numerically. What was novel about this work? First, dynamics on a strange attractor embedded in a high-dimensional space was essentially reduced to 1-dimensional dynamics. Second, the solutions found provided both a *qualitative description* and highly accurate *quantitative predictions* for the given PDE with the given boundary conditions and system parameter values.

By the end of the century, and struggling with the same computational limitations, Genta Kawahara and Shigeo Kida attempted to extend Nagata's work to periodic orbits. They avoided the large-matrix-problem altogether by using a direction set method to minimize the residual, i.e. the mismatch of the almost-periodic orbit. The down side was, that this method converges very slowly. Their computations ran for months on all desktop computers they could get their hands on. Early versions of their report were held back by the reviewers of the Journal of Fluid Mechanics, who finally relented when the relative residual had reached the magical number of 0.01. Thus, Kawahara and Kida became the first to compute a periodic orbit in a turbulent flow [109]. This computation was a herald of more exciting results to come, but also made clear that new numerical methods would have to be developed.

In the 2000s alternatives to the computation and decomposition of Jacobian

matrices were widely adopted in the community studying fluid dynamics as a dynamical system. These methods, collectively called *matrix-free*, or *Krylov subspace* methods require only a number of simulations of the flow along approximations of the invariant solution and the storage of the results. The number of simulations necessary is typically much smaller than the number of computational degrees of freedom, especially if the Reynolds number is not too high. Krylov methods for the computation of equilibria had been proposed as early as 1989 by Laurette Tuckerman [180], but their use increased dramatically after a Barcelona group worked out how to efficiently use this approach for the computation of periodic orbits [163]. A second development that accelerated this field of study was that of *edge tracking*, originally proposed by Itano and Toh [103] and successfully used, for instance, by Skufca, Yorke and Eckhardt in a 9-variable model [175]. This method exploits the fact that in many parallel shear flows, for instance, channel and pipe flows, the laminar flow is stable and thus, if there is any other kind of stable motion possible, there must exist an “edge” to its domain of attraction. Orbits inside this edge can be accurately traced and often converge on interesting invariant solutions.

The skepticism about dynamical systems-flavoured results on turbulence that pioneers like Nagata, Kawahara and Kida met with dissipated as increasingly convincing results appeared in the literature. Viswanath verified the results of Kawahara and Kida with great accuracy and was first to compute relative periodic orbit solutions [184]. Faisst and Eckhardt and Wedin and Kerswell independently presented a number of travelling wave solutions in pipe flow [69, 188] while Toh and Itano dissected Poiseuille flow [179]. Flow on a periodic domain was studied by van Veen, Kawahara and Kida [181]. Fuelling the developments further was the experimental observations of traces of travelling wave solutions in pipe flow by Björn Hof *cum suis* [98]. After all, fluid mechanics is an empirical field of study, and the relevance of the computational results, exciting as they were to the dynamicists, was not immediately clear to the larger physics and engineers.

By the end of the decade, new results were coming out every month. It is fair to say that anyone with a sufficiently accurate simulation code, studying sufficiently mild turbulence – at a Reynolds number in the hundreds, not thousands – could compose a Newton-Krylov solver and enter the game. In fact, several studies were launched in which equilibria and periodic orbits were computed semi-automatically to build up a catalog of “building blocks” or “tiles” of turbulence. Two results worth highlighting are those by Halcrow, Gibson, Cvitanović and Viswanath [96] on plane Couette flow, who include approximate connecting orbits – the outlines of the behaviour envisioned by Hopf – and those by Dan Lucas and Rich Kerswell on two-dimensional, spatially periodic flow [129].

How is it possible that the theory originally developed for low dimensional dynamical systems can work in the ∞ -dimensional PDE state spaces? For dissipative flows the number of unstable, expanding directions is often finite and even low-dimensional; perturbations along the ∞ of contracting directions heal themselves, and play only a minor role in cycle weights - hence the long-time dynamics is effectively finite dimensional. For a more precise statement, see Ginelli *et al.* [82].

remark 30.1

Although a general proof for three-dimensional Navier-Stokes flow is not available, it is widely assumed that turbulent dynamics, daunting as it may seem, is also finite-dimensional by grace of the viscous damping. If so, there is hope that a description in terms of solutions such as equilibria, periodic orbits and travelling waves, and a quasi-random trajectory careening from one to the other, is feasible and useful. In order to prove Hopf right, one has to compute such solutions - as many as needed to attain the desired accuracy for periodic orbit theory predictions [9].

All results described above are based on iterative methods, meaning that they all required some initial guess for the invariant solution of interest. Obtaining such an initial guess is a surprisingly tricky business. Some solutions spring from others at bifurcation points and can thus be tracked. Some live on the edge of the domain of attraction of others and can be identified using edge tracking. Some can be tracked by creating a homotopy, i.e. by continuously connecting the flow of interest to one with a tractable bifurcation diagram. For many, including the above results on Couette and pipe flow, no such gangway is available. In that case, the last resort is to start from a coarse approximation filtered from a turbulent time series. The convergence of iterative methods is then far from guaranteed. As we push the envelope, studying more and more turbulent flows and larger and larger domains, edging closer to the reproduction of experiments, finding a good starting point becomes more problematic. Two approaches that mitigate this problem are *globally convergent Newton iteration* and *variational methods*. The former were integrated with Krylov subspace methods by Viswanath [184]. The latter were developed by Lan and Cvitanović [51, 121, 122] and further developed by Farazmand [70].

Combining the best available methods, and using modern tools like GPU computing, various groups are tackling such challenging problems as boundary layers [113] and self-similar dynamics [182]. Little by little, the fluid dynamics community is starting to appreciate that the *dynamical* analysis of turbulent flows is now feasible, and provides a valuable complement to the traditional, *statistical* analysis.

A1.6 Gruppenpest

How many Tylenols should I take with this?... (never took group theory, still need to be convinced that there is any use to this beyond mind-numbing formalizations.)

— Fabian Waleffe, forced to read chapter 10.

If you are not fan of chapter 10 “Flips, slides and turns,” and its elaborations, you are not alone. Or, at least, you were not alone in the 1930s. That is when the articles by two young mathematical physicists, Eugene Wigner and Johann von Neumann [142], and Wigner’s 1931 Gruppentheorie [190] started *Die Gruppenpest* that plagues us to this very day.

question 10.1

According to John Baez [13], the American physicist John Slater, inventor of the ‘Slater determinant,’ is famous for having dismissed groups as unnecessary to physics. He wrote:

“It was at this point that Wigner, Hund, Heitler, and Weyl entered the picture with their ‘Gruppenpest:’ the pest of the group theory [actually, the correct translation is ‘the group plague’] ... The authors of the ‘Gruppenpest’ wrote papers which were incomprehensible to those like me who had not studied group theory... The practical consequences appeared to be negligible, but everyone felt that to be in the mainstream one had to learn about it. I had what I can only describe as a feeling of outrage at the turn which the subject had taken ... it was obvious that a great many other physicists were disgusted as I had been with the group-theoretical approach to the problem. As I heard later, there were remarks made such as ‘Slater has slain the ‘Gruppenpest’’. I believe that no other piece of work I have done was so universally popular.”

A. John Coleman writes in *Groups and Physics - Dogmatic Opinions of a Senior Citizen* [36]: “The mathematical elegance and profundity of Weyl’s book [Theory of Groups and QM] was somewhat traumatic for the English-speaking physics community. In the preface of the second edition in 1930, after a visit to the USA, Weyl wrote, “It has been rumored that the ‘group pest’ is gradually being cut out of quantum physics. This is certainly not true in so far as the rotation and Lorentz groups are concerned; ...” In the autobiography of J. C. Slater, published in 1975, the famous MIT physicist described the “feeling of outrage” he and other physicists felt at the incursion of group theory into physics at the hands of Wigner, Weyl et al. In 1935, when Condon and Shortley published their highly influential treatise on the “Theory of Atomic Spectra”, Slater was widely heralded as having “slain the Gruppenpest”. Pages 10 and 11 of Condon and Shortley’s treatise are fascinating reading in this context. They devote three paragraphs to the role of group theory in their book. First they say, “We manage to get along without it.” This is followed by a lovely anecdote. In 1928 Dirac gave a seminar, at the end of which Weyl protested that Dirac had said he would make no use of group theory but that in fact most of his arguments were applications of group theory. Dirac replied, “I said that I would obtain the results without previous knowledge of group theory!” Mackey, in the article referred to previously, argues that what Slater and Condon and Shortley did was to rename the generators of the Lie algebra of $SO(3)$ as “angular momenta” and create the feeling that what they were doing was physics and not esoteric mathematics.”

From [AIP Wigner interview](#): AIP: “In that circle of people you were working with in Berlin, was there much interest in group theory at this time?” WIGNER: “No. On the opposite. Schrödinger coined the expression, ‘Gruppenpest’ must be abolished.” “It is interesting, and representative of the relations between mathematics and physics, that Wigner’s paper was originally submitted to a Springer physics journal. It was rejected, and Wigner was seeking a physics journal that might take it when von Neumann told him not to worry, he would get it into the Annals of Mathematics. Wigner was happy to accept his offer [166].”

You would think it was all up from there for group theory. But no. In the

early 1970's, in the [nonexistent city of Bielefeld](#), writes M. du Sautoy [165]; “The Maoist movement decided that group theory was a reactionary subject of the old regime, and [...] demonstrations erupted outside of the maths department with protesters holding placards demanding ‘No more group theory’. [...] During one demonstration, the students scaled the outside of the building and scrawled ‘Group Theory Department’ on the wall.

A1.7 Death of the Old Quantum Theory

In 1913 Otto Stern and Max Theodor Felix von Laue went up for a walk up the Uetliberg. On the top they sat down and talked about physics. In particular they talked about the new atom model of Bohr. There and then they made the ‘Uetli Schwur:’ If that crazy model of Bohr turned out to be right, then they would leave physics. It did and they didn’t.

— A. Pais, *Inward Bound: of Matter and Forces in the Physical World*

One afternoon in May 1991, Dieter Wintgen is sitting in his office at the Niels Bohr Institute beaming with the unparalleled glee of a boy who has just committed a major mischief. The starting words of the manuscript he has just penned are

The failure of the Copenhagen School to obtain a reasonable ...

Wintgen was 34 years old at the time, a scruffy kind of guy, always wearing sandals and holed out jeans, the German flavor of a 90's left winger and mountain climber. He worked around the clock with his students Gregor Tanner and Klaus Richter to complete the work that Bohr himself would have loved to have seen done back in 1916: a ‘planetary’ calculation of the helium spectrum.

Never mind that the ‘Copenhagen School’ refers not to the old quantum theory, but to something else. The old quantum theory was no theory at all; it was a set of rules bringing some order to a set of phenomena which defied logic of deterministic theory. The electrons were supposed to describe planetary orbits around the nucleus; their wave aspects were yet to be discovered. The foundations seemed obscure, but Bohr’s answer for the once-ionized helium to hydrogen ratio was correct to five significant figures and hard to ignore. The old quantum theory marched on, until by 1924 it reached an impasse: the helium spectrum and the Zeeman effect were its death knell.

Since the late 1890's it had been known that the helium spectrum consists of the orthohelium and parahelium lines. In 1915 Bohr suggested that the two kinds of helium lines might be associated with two distinct shapes of orbits (a suggestion that turned out to be wrong). In 1916 he got Hans Kramers to work on the problem, and he wrote to Rutherford, “I have used all my spare time in the last months to make a serious attempt to solve the problem of ordinary helium

spectrum ... I think really that at last I have a clue to the problem.” To other colleagues he wrote that “the theory was worked out in the fall of 1916” and of having obtained a “partial agreement with the measurements.” Nevertheless, the Bohr-Sommerfeld theory, while by and large successful for hydrogen, was a disaster for neutral helium. Heroic efforts of the young generation, including Kramers and Heisenberg, were of no avail.

For a while Heisenberg thought that he had the ionization potential for helium, which he had obtained by a simple perturbative scheme. He wrote enthusiastic letters to Sommerfeld and was drawn into a collaboration with Max Born to compute the spectrum of helium using Born’s systematic perturbative scheme. To a first approximation, they reproduced the earlier calculations. The next level of corrections turned out to be larger than the computed effect. The concluding paragraph of Max Born’s classic “Vorlesungen über Atommechanik” from 1925 sums it up in a somber tone [24]:

(...) the systematic application of the principles of the quantum theory (...) gives results in agreement with experiment only in those cases where the motion of a single electron is considered; it fails even in the treatment of the motion of the two electrons in the helium atom.

This is not surprising, for the principles used are not really consistent. (...) A complete systematic transformation of the classical mechanics into a discontinuous mechanics is the goal towards which the quantum theory strives.

That year Heisenberg suffered a bout of hay fever, and the old quantum theory was dead. In 1926 he gave the first quantitative explanation of the helium spectrum. He used wave mechanics, electron spin and the Pauli exclusion principle, none of which belonged to the old quantum theory. As a result, planetary orbits of electrons were cast away for nearly half a century.

Why did Pauli and Heisenberg fail with the helium atom? It was not the fault of the old quantum mechanics, but rather it reflected their lack of understanding of the subtleties of classical mechanics. Today we know what they missed in 1913-24, the role of conjugate points (topological indices) along deterministic trajectories was not accounted for, and they had no idea of the importance of periodic orbits in nonintegrable systems.

Since then the calculation for helium using the methods of the old quantum mechanics has been fixed. Leopold and Percival [124] added the topological indices in 1980, and in 1991 Wintgen and collaborators [68, 191] understood the role of periodic orbits. Dieter had good reasons to gloat; while the rest of us were preparing to sharpen our pencils and supercomputers in order to approach the dreaded 3-body problem, they just went ahead and did it. What it took—and much else—is described in this book.

One is also free to ponder what quantum theory would look like today if all this was worked out in 1917. In 1994 Predrag Cvitanović gave a talk in Seattle about helium and cycle expansions to—inter alia—Hans Bethe, who loved it so much that

after the talk he pulled Predrag aside and they trotted over to Hans' secret place: the best lunch on campus (Business School). Predrag asked: "Would quantum mechanics look different if in 1917 Bohr and Kramers *et al.* figured out how to use the helium deterministic 3-body dynamics to quantize helium?"

Bethe was very annoyed. He responded with an exasperated look - in Bethe Deutschinglish (if you have ever talked to him, you can do the voice over yourself):

"It would not matter at all!"

Well, perhaps appendix [A39](#) proves him wrong...

Commentary

Remark A1.1. Notion of global foliations. For each paper cited in dynamical systems literature, there are many results that went into its development. As an example, take the notion of global foliations that we attribute to Smale. As far as we can trace the idea, it goes back to René Thom; local foliations were already used by Hadamard. Smale attended a seminar of Thom in 1958 or 1959. In that seminar Thom was explaining his notion of transversality. One of Thom’s disciples introduced Smale to Brazilian mathematician Peixoto. Peixoto (who had learned the results of the Andronov-Pontryagin school from Lefschetz) was the closest Smale had ever come until then to the Andronov-Pontryagin school. It was from Peixoto that Smale learned about structural stability, a notion that got him enthusiastic about dynamical systems, as it blended well with his topological background. It was from discussions with Peixoto that Smale got the problems in dynamical systems that lead him to his 1960 paper on Morse inequalities. The next year Smale published his result on the hyperbolic structure of the non-wandering set. Smale was not the first to consider a hyperbolic point, Poincaré had already done that; but Smale was the first to introduce a global hyperbolic structure. By 1960 Smale was already lecturing on the horseshoe as a structurally stable dynamical system with an infinity of periodic points and promoting his global viewpoint. (R. Mainieri)

Remark A1.2. A brief history of period doubling universality. Mitchell J. Feigenbaum discovered universality in one-dimensional iterative maps in August 1975. Following Feigenbaum’s functional formulation of the problem, in March 1976 Cvitanović derived, in collaboration with Feigenbaum, the equation $g(x) = \alpha g(g(x/\alpha))$ for the period doubling fixed point function (not a big step, it is the limit of Feigenbaum functional recursion sequence), which has since played a key role in the theory of transitions to turbulence. The first published report [71] on Feigenbaum’s discovery is dated August 1976 (Los Alamos Theoretical Division Annual Report 1975-1976, pp. 98-102, [read it here](#)). By that time the work had become widely known through many seminars Feigenbaum gave in US and Europe. His first paper, submitted to *Advances in Mathematics* in Nov 1976 was rejected. The second paper was submitted to *SIAM Journal of Applied Mathematics* in April 1977 and rejected in October 1977. Finally, J. Lebowitz published both papers [72, 73] without further referee pain (M. J. Feigenbaum, *J. Stat. Phys.* 19, 25 (1978) and 21, 6 (1979)).

A very informative 1976 review by May [134] describes what was known before Feigenbaum’s contribution. The geometric parameter convergence was first noted in 1958 by Myrberg [17, 164], and independently of Feigenbaum, by Grossmann and Thomae [87] in 1977 (without noting the universality of δ). The theory of period-doubling universal equations and scaling functions is developed in Kenway’s notes of Feigenbaum 1984 Edinburgh lectures [75] (trifle hard to track down). The elegant unstable manifold formulation of universality given in ChaosBook.org is due to Vul, Khanin, Sinai and Gol’dberg [83, 185, 186] in 1982. The most thorough exposition available is the Collet and Eckmann [37] monograph. For an introduction into renormalization theory that starts out with period doubling before moving on to Quantum Field Theory, see Gurau, Rivasseau and Sfondrini [88].

In 1978 Couillet and Tresser [39, 40] have formulated similar equations, in 1979 Derrida, Gervois and Pomeau [59] have extracted a great many metric universalities from the asymptotic regime, and in 1981 Daido [57] has introduced a different set of universal equations. Grassberger [84] has computed the Hausdorff dimension of the

asymptotic attractor. Following up on Grossmann and Thomae [87], Lorenz [128] and Daido [58] have found a universal ratio relating bifurcations and reverse bifurcations. If $f(x)$ is not quadratic around the maximum, the universal numbers will be different - see Vilela Mendés [183] and Hu and Mao [102] for their values. According to Kuramoto and Koga [118] such mappings can arise in chemical turbulence. Nonlinear oscillator; quadratic potential with damping and harmonic driving force exhibit cascades of period-doubling bifurcations [120, 138]. Refs. [26–28] compute solutions of the period-doubling fixed point equation using methods of Schöder and Abel, yielding what are so far the most accurate δ and α . See also Weisstein [189].

Since then the universal equations have been generalized to period n -tuplings [52, 53]; universal scaling functions for all winding numbers in circle maps constructed [54], and universality of the Hausdorff dimension of the critical staircase established [50]. A nice discussion of circle maps and their physical applications is given in refs. [14, 105, 106]. The universality theory for golden mean scalings is developed in refs. [74, 143, 152, 169].

The theory would have remained a curiosity, were it not for the beautiful experiment by Libchaber and Maurer [133], and many others that followed. Crucial insights came from Collet and Eckmann [37] and Collet, Eckmann and Koch [38] who explained how the dynamics of dissipative system (such as a viscous fluid) can become 1-dimensional. The experimental and theoretical developments up to 1990's are summarized in reprint collections by Cvitanović [44] and Hao [97]. We also recommend Hu [101], Crutchfield, Farmer and Huberman [42], Eckmann [66] and Ott [144]. The period-doubling route to turbulence that is by no means the only way to get there; see Eckmann [66] discussion of other routes to chaos.

Remark A1.3. Should one attach names to equations? .

Q : Name the 2nd person [who invented General Relativity](#)?

A : Who remembers?

—Prof. Dr. Dr. h.c. Gatto Nero, Grande Ufficiale dell'Ordine della Stella della Notorietà Italiana

By 1979 mathematicians understood that the numerical methods used by Feigenbaum and Cvitanović to solve the universal equations were in fact convergent. They did what they do; they attached various names to the equations, they changed letters around. The re-lettering did not stick, but the renamings did.

Feigenbaum [71] discovered and formulated period-doubling universality in 1975: you can read about it and find his 1976 report by [clicking here](#) and [here](#). In 1981 Lanford [123] satisfied himself that the iterative method Feigenbaum and Cvitanović used and knew was contracting was indeed contracting. Lanford refers only to the Feigenbaum paper [72]. [Coullet and Tresser](#) [39, 40] refer to the Feigenbaum paper [72].

In 1995 Lyubich [131, 132] rechristened the equations to “Feigenbaum-Coullet-Tresser,” omitting Cvitanović (the first to formulate the period-doubling fixed point equation), and adding Coullet and Tresser (who rediscovered it a couple of years later). These are all very fine physicists / mathematicians, creative and crazy as bats. But why rename an equation that was widely known and publicized well before 1978? Is there something essential that is missing in the 1976 formulation?

We asked Lyubich why? He wrote back: “In 1990s, I talked to both Feigenbaum and Tresser, and my conclusion was that Coullet-Tresser discovered the phenomenon inde-

pendently, though slightly later. Also, they seemed to recognize better importance of the dynamical universality (while Feigenbaum focused more on the parameter phenomenon). I felt that Couillet-Tresser did not receive a proper credit for their insights, so I attached all three names to the phenomenon.” That’s sweet. Turns out Feigenbaum and Cvitanović invented but did not recognize “importance of the dynamical universality”, whatever that might mean. While we are at it, why not credit the person who actually wrote the fixed point equation first? Or he’s just dog meat?

People reinvent stuff all the time. For example, Myrheim and Cvitanović [52, 53] generalized period doubling to [infinity of renormalizations](#) in the complex plane, but once they were told that Golberg, Sinai and Khanin [83] did it first (for period tripling), they gave credit to them, even though both groups discovered the phenomenon independently in 1983.

Why attach names to equations anyway? Pretty soon the attribution problems will sort themselves out by themselves - [heart attacks](#) and homicidal Atlanta drivers running down cyclists will take care of that.

Remark A1.4. Levels of ergodicity. In the mid 1970’s A. Katok and Ya.B. Pesin tried to use geometry to establish positive Lyapunov exponents. A. Katok and J.-M. Strelcyn carried out the program and developed a theory of general dynamical systems with singularities. They studied uniformly hyperbolic systems (as strong as Anosov’s), but with sets of singularities. Under iterations a dense set of points hits the singularities. Even more important are the points that never hit the singularity set. In order to establish some control over how they approach the set, one looks at trajectories that approach the set by some given ϵ^n , or faster.

Ya.G. Sinai, L. Bunimovich and N.I. Chernov studied the geometry of billiards in a very detailed way. A. Katok and Ya.B. Pesin’s idea was much more robust: look at the discontinuity set, take an ϵ neighborhood around it. Given that the Lebesgue measure is ϵ^α and the stability grows not faster than (distance) ^{n} . A. Katok and J.-M. Strelcyn proved that the Lyapunov exponent is non-zero.

In mid 1980’s Ya.B. Pesin studied the dissipative case. Now the problem has no invariant Lebesgue measure. Assuming uniform hyperbolicity, with singularities, and tying together Lebesgue measure and discontinuities, and given that the stability grows not faster than (distance) ^{n} , Ya.B. Pesin proved that the Lyapunov exponent is non-zero, and that SRB measure exists. He also proved that the Lorenz, Lozi and Byelikh attractors satisfy these conditions.

In the systems that are uniformly hyperbolic, all trouble is in differentials. For the Hénon attractor, already the differentials are nonhyperbolic. The points do not separate uniformly, but the analogue of the singularity set can be obtained by excising the regions that do not separate. Hence there are 3 levels of ergodic systems:

1. Anosov flow
2. Anosov flow + singularity set: For the Hamiltonian systems the general case is studied by A. Katok and J.-M. Strelcyn, and the billiards case by Ya.G. Sinai and L. Bunimovich. The dissipative case is studied by Ya.B. Pesin.
3. Hénon case: The first proof was given by M. Benedicks and L. Carleson [18–20]. A more readable proof is given in M. Benedicks and L.-S. Young [21].

Remark A1.5. Is the geometry of nature fractal? By 1983 some physicists were starting to learn that there is a thing called “chaos” [125], a thing stressful, nasty and hard

to understand (see this book), so they tried to bypass this whole bit of unpleasantness by getting instead an easy, diagnostic number out of it [108]. They were told that Hausdorff dimension is the way to go. Dimension of the canonical $1/3$'s Cantor set can be explained to a school child, so they tried it out for size on many low-dimensional chaotic attractors, and some crazy high-dimensional ones as well. Our all-time favorite (beyond the 'Untitled 5' of figure 1.5) was the claim that the dimension of climate is 3.1 (remember, the policy of ChaosBook is not to pump up citation numbers for silly or plainly wrong papers): In *Deterministic chaos: the science and the fiction*, David Ruelle [160] comments: "[...] one should not believe dimension estimates that are not well below $2 \log_{10} N$. [Authors of ...] claim to find a dimension 3.1 for a 'climatic attractor' with $N = 500$ data points. [...] The 'dimensions' of the order 6 that are obtained are very close to the upper bound $2 \log_{10} N$ permitted by the Grassberger-Procaccia algorithm [85] (N is the length of the time series used, of the order of 103). The 'proof' that one has low dimensional dynamics is therefore inconclusive, and the suspicion is that the time evolutions under discussion do not correspond to low-dimensional dynamics. [...] Readers of *The Ultimate Hitchhiker's Guide to the Galaxy*, that masterpiece of British literature by D. Adams [1], know that a huge supercomputer has answered 'the great problem of life, the universe, and everything'. The answer obtained after many years of computation is 42. Unfortunately, one does not know to what precise question this is the answer, and what to make of it. It think that what happened is this. The supercomputer took a very long time series describing all it knew about 'life, the universe, and everything' and proceeded to compute the correlation dimension of the corresponding dynamics, using the Grassberger-Procaccia algorithm. This time series had a length N somewhat larger than 10^{21} . And you can imagine what happened. After many years of computation the answer came: dimension is approximately $2 \log_{10} \approx 42$." In 1998 Avnir, Biham, Lidar, and Malcai [12] explored how much support for the fractal self-similarity hypothesis was there, actually. They found that "the majority of the data that was interpreted in terms of fractality in the surveyed Physical Review journals does not seem to be linked (at least in an obvious way) to existing models and, in fact, does not have theoretical backing. Most of the data represent results from nonequilibrium processes. The common situation is this: An experimentalist performs a resolution analysis and finds a limited-range power law with a value of D smaller than the embedding dimension. Without necessarily resorting to special underlying mechanistic arguments, the experimentalist then often chooses to label the object for which she or he finds this power law a 'fractal'. This is the fractal geometry of nature." Their plot says it all: the number of decades (factors of 10) spanned by experimentally derived scaling exponents peaks at 10 (i.e., one decade).

Remark A1.6. Einstein did it? The first hint that chaos is afoot in quantum mechanics was given in a note by A. Einstein [67]. The total discussion is a one sentence remark. Einstein being Einstein, this one sentence has been deemed sufficient to give him the credit for being the pioneer of quantum chaos [94, 178]. We asked about the paper two people from that era, Sir Rudolf Peierls and Abraham Pais; neither had any recollection of the 1917 article. However, Theo Geisel has unearthed a reference that shows that in early 20s Born did have a study group meeting in his house that studied Poincaré's *Mécanique Céleste* [149]. In 1954 Fritz Reiche, who had previously followed Einstein as professor of physics in Breslau (now Wrocław, Poland), pointed out to J.B. Keller that Keller's geometrical semiclassical quantization was anticipated by the long forgotten paper by A. Einstein [67]. In this way an important paper written by the physicist who at the time was the president of German Physical Society, and the most famous scientist of his time,

came to be referred to for the first time by Keller [112], 41 years later. But before Ian Percival included the topological phase, and Wintgen and students [68, 191] recycled the Helium atom, knowing *Mécanique Céleste* was not enough to complete Bohr's original program.

Remark A1.7. Berry-Keating conjecture. A very appealing proposal in the context of semiclassical quantization is due to M. Berry and J. Keating [23]. The idea is to improve cycle expansions by imposing unitarity as a functional equation ansatz. The cycle expansions that they use are the same as the original ones described above [8], but the philosophy is quite different; the claim is that the optimal estimate for low eigenvalues of deterministically chaotic quantum systems is obtained by taking the real part of the cycle expansion of the semiclassical zeta function, cut off at the appropriate cycle length. M. Sieber, G. Tanner and D. Wintgen, and P. Dahlqvist find that their numerical results support this claim; F. Christiansen and P. Cvitanović do not find any evidence in their numerical results. The usual Riemann-Siegel formulas exploit the self-duality of the Riemann and other zeta functions, but there is no evidence of such symmetry for generic Hamiltonian flows. Also from the point of hyperbolic dynamics discussed above, proposal in its current form belongs to the category of crude cycle expansions; the cycles are cut off by a single external criterion, such as the maximal cycle time, with no regard for the topology and the curvature corrections. While the functional equation conjecture is not in its final form yet, it is very intriguing and fruitful research inspiration.

The real life challenge are generic dynamical flows, which fit neither of extreme idealized settings, Smale horseshoe on one end, and the Riemann zeta function on the other.

Remark A1.8. Sources. The tale of appendix A1.7, aside from a few personal recollections, is in large part lifted from Abraham Pais' accounts of the demise of the old quantum theory [146, 147], as well as Jammer's account [104]. In August 1994 Dieter Wintgen died in a climbing accident in the Swiss Alps. Remark A1.4 is based on Ya.B. Pesin's comments.

References

- [1] D. Adams, *The Ultimate Hitchhiker's Guide to the Galaxy* (Random House, New York, 1979).
- [2] R. C. Adler, A. C. Konheim, and M. H. McAndrew, "Topological entropy", *Trans. Amer. Math. Soc.* **114**, 309–319 (1965).
- [3] R. L. Adler and R. Palais, "Homeomorphic conjugacy of automorphisms on the torus", *Proc. Amer. Math. Soc.* **16**, 1222–1222 (1965).
- [4] E. L. Allgower and K. Georg, *Introduction to Numerical Continuation Methods* (SIAM, 2003).
- [5] D. V. Anosov and Y. G. Sinai, "Some smooth ergodic systems", *Russ. Math. Surv.* **22**, 103–167 (1967).
- [6] M. Artin and B. Mazur, "On periodic points", *Ann. Math.* **81**, 82–99 (1965).
- [7] R. Artuso, "Diffusive dynamics and periodic orbits of dynamic systems", *Phys. Lett. A* **160**, 528–530 (1991).

- [8] R. Artuso, E. Aurell, and P. Cvitanović, “Recycling of strange sets: I. Cycle expansions”, *Nonlinearity* **3**, 325–359 (1990).
- [9] R. Artuso, E. Aurell, and P. Cvitanović, “Recycling of strange sets: II. Applications”, *Nonlinearity* **3**, 361–386 (1990).
- [10] N. Aubry, P. Holmes, J. L. Lumley, and E. Stone, “The dynamics of coherent structures in the wall region of turbulent boundary layer”, *J. Fluid Mech.* **192**, 115–173 (1988).
- [11] D. Auerbach, P. Cvitanović, J.-P. Eckmann, G. Gunaratne, and I. Procaccia, “Exploring chaotic motion through periodic orbits”, *Phys. Rev. Lett.* **58**, 2387–2389 (1987).
- [12] D. Avnir, O. Biham, D. Lidar, and O. Malcai, “Is the geometry of nature fractal?”, *Science* **279**, 39–40 (1998).
- [13] J. Baez, *This Week’s Finds in Mathematical Physics* Week 236, 2006.
- [14] P. Bak, T. Bohr, and M. H. Jensen, “Mode-locking and the transition to chaos in dissipative systems”, *Physica Scripta* **T9**, 50–58 (1985).
- [15] V. Baladi, *Positive Transfer Operators and Decay of Correlations* (World Scientific, Singapore, 2000).
- [16] J. Barrow-Green, *Poincaré and the Three Body Problem* (Amer. Math. Soc., Providence R.I., 1997).
- [17] M. W. Beims and J. A. C. Gallas, “Accumulation points in nonlinear parameter lattices”, *Physica A* **238**, 225–244 (1997).
- [18] M. Benedicks and L. Carleson, “On iterations of $1 - ax^2$ on $(-1, 1)$ ”, *Ann. Math.* **122**, 1–25 (1985).
- [19] M. Benedicks and L. Carleson, On the Hénon attractor, in IXth Int. Congr. on Mathematical Physics, edited by B. Simon, A. Truman, and M. Davies (1989), pp. 498–500.
- [20] M. Benedicks and L. Carleson, “The dynamics of the Hénon map”, *Ann. Math.* **133**, 73 (1991).
- [21] M. Benedicks and L.-S. Young, “Absolutely continuous invariant measures and random perturbations for certain one-dimensional map”, *Ergod. Theor. Dynam. Syst.* **12**, 13–37 (1992).
- [22] M. V. Berry, “Martin Gutzwiller and his periodic orbits”, *Commun. Swiss Phys. Soc.* **37**, 34–38 (2012).
- [23] M. V. Berry and J. P. Keating, “A rule for quantizing chaos”, *J. Phys. A* **23**, 4839–4849 (1990).
- [24] M. Born, *Vorlesungen über Atommechanik*, English Translation: The Mechanics of the Atom (Ungar Publishing, New York, 1927).
- [25] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms* (Springer, New York, 1975).
- [26] K. Briggs, “A precise calculation of the Feigenbaum constants”, *Mathematics of Computation* **57**, 435–439 (1991).

- [27] K. M. Briggs, T. W. Dixon, and G. Szekeres, “Analytic solutions of the Cvitanović-Feigenbaum and Feigenbaum-Kadanoff-Shenker equations”, *Int. J. Bifur. Chaos* **8**, 347–357 (1998).
- [28] K. M. Briggs, G. R. W. Quispel, and C. J. Thompson, “Eigenvalues for Mandelsets”, *J. Phys. A* **24**, 3363–3368 (1991).
- [29] H. Bruns, “Über die Integrale des Vielkörper-Problems”, *Acta Math.* **11**, 25–96 (1887).
- [30] M. L. Cartwright and J. E. Littlewood, “On non-linear differential equations of the second order”, *J. London Math. Soc.* **20**, 180–189 (1945).
- [31] F. Christiansen and P. Cvitanović, “Periodic orbit quantization of the anisotropic Kepler problem”, *Chaos* **2**, 61–69 (1992).
- [32] F. Christiansen, P. Cvitanović, and V. Putkaradze, “Hopf’s last hope: Spatiotemporal chaos in terms of unstable recurrent patterns”, *Nonlinearity* **10**, 55–70 (1997).
- [33] F. Christiansen, P. Cvitanović, and H. H. Rugh, “The spectrum of the period-doubling operator in terms of cycles”, *J. Phys. A* **23**, L713S–L717S (1990).
- [34] K. A. Cliffe, *ENTWIFE (release 6.3) reference manual*, tech. rep. (Oxford, UK: Harwell Laboratory, 1996).
- [35] K. A. Cliffe and T. Mullin, “A numerical and experimental study of anomalous modes in the Taylor experiment”, *J. Fluid Mech.* **153**, 243–258 (1985).
- [36] A. J. Coleman, “Groups and physics – Dogmatic opinions of a senior citizen”, *Notices Amer. Math. Soc.* **44**, 8–17 (1997).
- [37] P. Collet and J.-P. Eckman, *Iterated Maps on the Interval as Dynamical Systems* (Birkhäuser, Boston, 2009).
- [38] P. Collet, J.-P. Eckmann, and H. Koch, “Period doubling bifurcations for families of maps on R_n ”, *J. Stat. Phys.* **25**, 1–14 (1981).
- [39] P. Coullet and C. Tresser, “Itérations d’endomorphismes et groupe de renormalisation”, *J. Phys. Colloques C5* **39**, 25–28 (1978).
- [40] P. Coullet and C. Tresser, “Iterations of endomorphisms and renormalization group”, *C. R. Acad. Sc. Paris A* **287**, 577–581 (1978).
- [41] G. Cristadoro, “Fractal diffusion coefficient from dynamical zeta functions”, *J. Phys. A* **39**, L151 (2006).
- [42] J. P. Crutchfield, J. D. Farmer, and B. A. Huberman, “Fluctuations and simple chaotic dynamics”, *Phys. Rep.* **92**, 45–82 (1982).
- [43] P. Cvitanović, “Invariant measurement of strange sets in terms of cycles”, *Phys. Rev. Lett.* **61**, 2729–2732 (1988).
- [44] P. Cvitanović, *Universality in Chaos*, 2nd ed. (Adam Hilger, Bristol, 1989).
- [45] P. Cvitanović and B. Eckhardt, “Periodic orbit quantization of chaotic systems”, *Phys. Rev. Lett.* **63**, 823–826 (1989).
- [46] P. Cvitanović and B. Eckhardt, “Periodic orbit expansions for classical smooth flows”, *J. Phys. A* **24**, L237 (1991).

- [47] P. Cvitanović and B. Eckhardt, “Symmetry decomposition of chaotic dynamics”, *Nonlinearity* **6**, 277–311 (1993).
- [48] P. Cvitanović, J.-P. Eckmann, and P. Gaspard, “Transport properties of the Lorentz gas in terms of periodic orbits”, *Chaos Solit. Fract.* **6**, 113–120 (1995).
- [49] P. Cvitanović, G. H. Gunaratne, and I. Procaccia, “Topological and metric properties of Hénon-type strange attractors”, *Phys. Rev. A* **38**, 1503–1520 (1988).
- [50] P. Cvitanović, G. H. Gunaratne, and M. J. Vinson, “On the mode-locking universality for critical circle maps”, *Nonlinearity* **3**, 873–885 (1990).
- [51] P. Cvitanović and Y. Lan, Turbulent fields and their recurrences, in *Correlations and Fluctuations in QCD : Proceedings of 10. International Workshop on Multiparticle Production*, edited by N. Antoniou (2003), pp. 313–325.
- [52] P. Cvitanović and J. Myrheim, “Universality for period n -tuplings in complex mappings”, *Phys. Lett. A* **94**, 329–333 (1983).
- [53] P. Cvitanović and J. Myrheim, “Complex universality”, *Commun. Math. Phys.* **121**, 225–254 (1989).
- [54] P. Cvitanović, B. Shraiman, and B. Söderberg, “Scaling laws for mode lockings in circle maps”, *Phys. Scr.* **32**, 263–270 (1985).
- [55] P. Dahlqvist, “Determination of resonance spectra for bound chaotic systems”, *J. Phys. A* **27**, 763–785 (1994).
- [56] P. Dahlqvist and G. Russberg, “Periodic orbit quantization of bound chaotic systems”, *J. Phys. A* **24**, 4763–4778 (1991).
- [57] H. Daido, “Universal relation of a band-splitting sequence to a preceding period-doubling one”, *Phys. Lett. A* **86**, 259–262 (1981).
- [58] H. Daido, “Period-doubling bifurcations and associated universal properties including parameter dependence”, *Progr. Theor. Phys.* **67**, 1698–1723 (1982).
- [59] B. Derrida, A. Grevois, and Y. Pomeau, “Universal metric properties of bifurcations of endomorphisms”, *J. Phys. A* **12**, 269–296 (1979).
- [60] C. P. Dettmann and G. P. Morriss, “Stability ordering; strong field Lorentz gas”, *Phys. Rev. Lett.* **78**, 4201–4204 (1997).
- [61] F. Diacu and P. Holmes, *Celestial Encounters: The Origins of Chaos and Stability* (Princeton Univ. Press, Princeton, NJ, 1996).
- [62] E. J. Doedel, “AUTO: a program for the automatic bifurcation analysis for autonomous systems”, *Congressus Numerantium* **30**, 265–284 (1981).
- [63] E. J. Doedel, A. R. Champneys, T. F. Fairgrieve, Y. A. Kuznetsov, B. Sandstede, and X. Wang, *AUTO: Continuation and Bifurcation Software for Ordinary Differential Equations*, tech. rep. (Computational Mathematics and Visualization Laboratory, Concordia Univ., 2007).
- [64] B. Eckhardt, “Fractal properties of scattering singularities”, *J. Phys. A* **20**, 5971–5979 (1987).

- [65] B. Eckhardt, G. Russberg, P. Cvitanović, P. E. Rosenqvist, and P. Scherer, “Pinball scattering”, in *Quantum Chaos: Between Order and Disorder*, edited by G. Casati and B. Chirikov (Cambridge Univ. Press, Cambridge UK, 1995).
- [66] J.-P. Eckmann, “Roads to turbulence in dissipative dynamical systems”, *Rev. Mod. Phys.* **53**, 643 (1981).
- [67] A. Einstein, “On the quantum theorem of Sommerfeld and Epstein”, in *The Collected Papers of Albert Einstein: The Berlin Years: Writings (1914–1917)*, Vol. 6, E. Schucking, English translation of “Zum Quantensatz von Sommerfeld und Epstein,” *Verh. Deutsch. Phys. Ges.* 19, 82 (Princeton Univ. Press, Princeton NJ, 1917), p. 443.
- [68] G. S. Ezra, K. Richter, G. Tanner, and D. Wintgen, “Semiclassical cycle expansion for the helium atom”, *J. Phys. B* **24**, L413–L420 (1991).
- [69] H. Faisst and B. Eckhardt, “Traveling waves in pipe flow”, *Phys. Rev. Lett.* **91**, 224502 (2003).
- [70] M. Farazmand, “An adjoint-based approach for finding invariant solutions of Navier-Stokes equations”, *J. Fluid Mech.* **795**, 278–312 (2016).
- [71] M. J. Feigenbaum, *Universality in complex discrete dynamics*, Los Alamos Theoretical Division Annual Report 1975–1976, 1976.
- [72] M. J. Feigenbaum, “Quantitative universality for a class of nonlinear transformations”, *J. Stat. Phys.* **19**, 25–52 (1978).
- [73] M. J. Feigenbaum, “The universal metric properties of nonlinear transformations”, *J. Stat. Phys.* **21**, Reprinted in ref. [44], 669–706 (1979).
- [74] M. J. Feigenbaum, L. P. Kadanoff, and S. J. Shenker, “Quasiperiodicity in dissipative systems: A renormalization group analysis”, *Physica D* **5**, 370–386 (1982).
- [75] M. J. Feigenbaum and R. D. Kenway, The onset of chaos, in *Statistical and Particle Physics: Common Problems and Techniques*. Proc. 26th Scottish Universities Summer School in Physics, edited by K. C. Bowler and A. J. McKane (1984), pp. 1–100.
- [76] C. Foias, B. Nicolaenko, G. R. Sell, and R. Témam, “Inertial manifolds for the Kuramoto-Sivashinsky equation and an estimate of their lowest dimension”, *J. Math. Pure Appl.* **67**, 197–226 (1988).
- [77] D. Fried, “Meromorphic zeta functions for analytic flows”, *Commun. Math. Phys.* **174**, 161–190 (1995).
- [78] P. Gaspard, *Chaos, Scattering and Statistical Mechanics* (Cambridge Univ. Press, Cambridge UK, 1998).
- [79] P. Gaspard and S. A. Rice, “Exact quantization of the scattering from a classically chaotic repeller”, *J. Chem. Phys.* **90**, 2255–2262 (1989).
- [80] P. Gaspard and S. A. Rice, “Scattering from a classically chaotic repeller”, *J. Chem. Phys.* **90**, 2225–2241 (1989).
- [81] P. Gaspard and S. A. Rice, “Semiclassical quantization of the scattering from a classically chaotic repeller”, *J. Chem. Phys.* **90**, 2242–2254 (1989).

- [82] F. Ginelli, P. Poggi, A. Turchi, H. Chaté, R. Livi, and A. Politi, “Characterizing dynamics with covariant Lyapunov vectors”, *Phys. Rev. Lett.* **99**, 130601 (2007).
- [83] A. I. Gol’berg, Y. G. Sinai, and K. M. Khanin, “Universal properties for sequences of bifurcations of period 3”, *Russ. Math. Surv.* **38**, 187–188 (1983).
- [84] P. Grassberger, “On the Hausdorff dimension of fractal attractors”, *J. Stat. Phys.* **26**, 173–179 (1981).
- [85] P. Grassberger and I. Procaccia, “Characterization of strange attractors”, *Phys. Rev. Lett.* **50**, 346–349 (1983).
- [86] C. Grebogi, E. Ott, and J. A. Yorke, “Unstable periodic orbits and the dimensions of multifractal chaotic attractors”, *Phys. Rev. A* **37**, 1711–1724 (1988).
- [87] S. Grossmann and S. Thomae, “Invariant distributions and stationary correlation functions of one-dimensional discrete processes”, *Z. Naturf. A* **32**, 1353–1363 (1977).
- [88] R. Gurau, V. Rivasseau, and A. Sfondrini, *Renormalization: An advanced overview*, 2014.
- [89] M. C. Gutzwiller, “Phase-integral approximation in momentum space and the bound states of an atom”, *J. Math. Phys.* **8**, 1979–2000 (1967).
- [90] M. C. Gutzwiller, “Phase-integral approximation in momentum space and the bound states of an atom. II”, *J. Math. Phys.* **10**, 1004–1020 (1969).
- [91] M. C. Gutzwiller, “Energy spectrum according to classical mechanics”, *J. Math. Phys.* **11**, 1791–1806 (1970).
- [92] M. C. Gutzwiller, “Periodic orbits and classical quantization conditions”, *J. Math. Phys.* **12**, 343–358 (1971).
- [93] M. C. Gutzwiller, “The quantization of a classically ergodic system”, *Physica D* **5**, 183–207 (1982).
- [94] M. C. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer, New York, 1990).
- [95] J. S. Hadamard, *Sur le Billard Non Euclidien* (Impr. Gauthier-Villars, 1898).
- [96] J. Halcrow, J. F. Gibson, P. Cvitanović, and D. Viswanath, “Heteroclinic connections in plane Couette flow”, *J. Fluid Mech.* **621**, 365–376 (2009).
- [97] B.-L. Hao, *Chaos II* (World Scientific, Singapore, 1990).
- [98] B. Hof, C. W. H. van Doorne, J. Westerweel, F. T. M. Nieuwstadt, H. Faisst, B. Eckhardt, H. Wedin, R. R. Kerswell, and F. Waleffe, “Experimental observation of nonlinear traveling waves in turbulent pipe flow”, *Science* **305**, 1594–1598 (2004).
- [99] P. Holmes, J. L. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge Univ. Press, Cambridge UK, 1996).
- [100] E. Hopf, “A mathematical example displaying features of turbulence”, *Commun. Pure Appl. Math.* **1**, 303–322 (1948).

- [101] B. Hu, “Introduction to real-space renormalization-group methods in critical and chaotic phenomena”, *Phys. Rep.* **91**, 233–295 (1982).
- [102] B. Hu and J. M. Mao, “Period doubling: Universality and critical-point order”, *Phys. Rev. A* **25**, 3259–3261 (1982).
- [103] T. Itano and S. Toh, “The dynamics of bursting process in wall turbulence”, *J. Phys. Soc. Japan* **70**, 703–716 (2001).
- [104] M. Jammer, *The Conceptual Development of Quantum Mechanics* (McGraw-Hill, New York, 1966).
- [105] M. H. Jensen, P. Bak, and T. Bohr, “Complete devil’s staircase fractal dimension and universality of mode-locking structure in the circle map”, *Phys. Rev. Lett.* **50**, 1637–1639 (1983).
- [106] M. H. Jensen, P. Bak, and T. Bohr, “Transition to chaos by interaction of resonances in dissipative systems. I. Circle maps”, *Phys. Rev. A* **30**, 1960–1969 (1984).
- [107] L. Kadanoff and C. Tang, “Escape rate from strange repellers”, *Proc. Natl. Acad. Sci. USA* **81**, 1276–1279 (1984).
- [108] L. P. Kadanoff, “Fractals: where’s the physics?”, *Phys. Today* **39**, 6 (1986).
- [109] G. Kawahara and S. Kida, “Periodic motion embedded in plane Couette turbulence: Regeneration cycle and burst”, *J. Fluid Mech.* **449**, 291–300 (2001).
- [110] E. Kazantsev, “Unstable periodic orbits and attractor of the barotropic ocean model”, *Nonlin. Proc. Geophys.* **5**, 193 (1998).
- [111] J. P. Keating, “Resummation and the turning points of zeta functions”, in *Classical, Semiclassical and Quantum Dynamics in Atoms*, edited by B. Eckhardt and H. Friedrich (Springer, Berlin, 1997), pp. 83–93.
- [112] J. B. Keller, “Corrected Bohr–Sommerfeld quantum conditions for non-separable systems”, *Ann. Phys. (N. Y.)* **4**, 180–188 (1958).
- [113] T. Khapko, T. Kreilos, P. Schlatter, Y. Duguet, B. Eckhardt, and D. S. Henningson, “Edge states as mediators of bypass transition in boundary-layer flows”, *J. Fluid Mech.* **801**, R2 (2016).
- [114] A. N. Kolmogorov, “A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces”, *Dokl. Akad. Nauk SSSR* **119**, 861–864 (1958).
- [115] A. N. Kolmogorov, “Entropy per unit time as a metric invariant of automorphisms”, *Dokl. Akad. Nauk SSSR* **124**, 754–755 (1959).
- [116] B. O. Koopman, “Hamiltonian systems and transformations in Hilbert space”, *Proc. Natl. Acad. Sci. USA* **17**, 315 (1931).
- [117] T. Kreilos and B. Eckhardt, “Periodic orbits near onset of chaos in plane Couette flow”, *Chaos* **22**, 047505 (2012).
- [118] Y. Kuramoto and S. Koga, “Anomalous period-doubling bifurcations leading to chemical turbulence”, *Phys. Lett. A* **92**, 1–4 (1982).

- [119] Y. Kuramoto and T. Tsuzuki, “Persistent propagation of concentration waves in dissipative media far from thermal equilibrium”, *Progr. Theor. Phys.* **55**, 365–369 (1976).
- [120] A. Y. Kuznetsova, C. K. A. P. Kuznetsov, and E. Mosekilde, “Catastrophe theoretic classification of nonlinear oscillators”, *Int. J. Bifur. Chaos* **12**, 1241–1266 (2004).
- [121] Y. Lan, *Dynamical Systems Approach to 1 – d Spatiotemporal Chaos – A Cyclist’s View*, PhD thesis (School of Physics, Georgia Inst. of Technology, Atlanta, 2004).
- [122] Y. Lan and P. Cvitanović, “Variational method for finding periodic orbits in a general flow”, *Phys. Rev. E* **69**, 016217 (2004).
- [123] O. E. Lanford, “A computer-assisted proof of the Feigenbaum conjectures”, *Bull. Amer. Math. Soc.* **6**, 427–434 (1982).
- [124] J. G. Leopold and I. Percival, “The semiclassical two-electron atom and the old quantum theory”, *J. Phys. B* **13**, 1037 (1980).
- [125] T.-Y. Li and J. A. Yorke, “Period three implies chaos”, *Amer. Math. Monthly* **82**, 985–992 (1975).
- [126] V. López, P. Boyland, M. T. Heath, and R. D. Moser, “Relative periodic solutions of the Complex Ginzburg-Landau equation”, *SIAM J. Appl. Dyn. Syst.* **4**, 1042 (2006).
- [127] E. N. Lorenz, “Deterministic nonperiodic flow”, *J. Atmos. Sci.* **20**, 130–141 (1963).
- [128] E. N. Lorenz, “Noisy periodicity and reverse bifurcation”, *Ann. New York Acad. Sci.* **357**, 282–291 (1980).
- [129] D. Lucas and R. R. Kerswell, “Recurrent flow analysis in spatiotemporally chaotic 2-dimensional Kolmogorov flow”, *Phys. Fluids* **27**, 518–554 (2015).
- [130] A. Lyapunov, “Problème général de la stabilité du mouvement”, *Ann. Math. Studies* **17**, Russian original Kharkow, 1892, 531–534 (1977).
- [131] M. Lyubich, “Renormalization ideas in conformal dynamics”, *Current Developments in Mathematics* **1995**, 155–190 (1995).
- [132] M. Lyubich, “Feigenbaum-Couillet-Tresser universality and Milnor’s hairiness conjecture”, *Ann. Math.* **149**, 319–420 (1999).
- [133] J. Maurer and A. Libchaber, “Effect of the Prandtl number on the onset of turbulence in liquid 4He”, *J. Physique Lett.* **41**, 515–518 (1980).
- [134] R. M. May, “Simple mathematical models with very complicated dynamics”, *Nature* **261**, 459–467 (1976).
- [135] H. P. McKean, “Selberg’s trace formula as applied to a compact Riemann surface”, *Commun. Pure Appl. Math.* **25**, 225–246 (1972).
- [136] N. Metropolis, M. L. Stein, and P. R. Stein, “On finite limit sets for transformations on the unit interval”, *J. Combin. Theory* **15**, 25–44 (1973).

- [137] J. Milnor and W. Thurston, “On iterated maps of the interval”, in *Dynamical Systems*, edited by J. C. Alexander (Springer, New York, 1988), pp. 465–563.
- [138] F. C. Moon, *Chaotic Vibrations: An Introduction for Applied Scientists and Engineers* (Wiley, New York, 1987).
- [139] D. W. Moore and E. A. Spiegel, “A thermally excited nonlinear oscillator”, *Astrophys. J.* **143**, 871–887 (1966).
- [140] M. Nagata, “Three-dimensional finite-amplitude solutions in plane Couette flow: bifurcation from infinity.” *J. Fluid Mech.* **217**, 519–527 (1990).
- [141] J. von Neumann, “Zusätze zur Arbeit “Zur Operatorenmethode in der klassischen Mechanik”. (German) [Additions to the work “On operator methods in classical mechanics”]”, *Ann. Math.* **33**, 789–791 (1932).
- [142] J. von Neumann and E. P. Wigner, “Über merkwürdige diskrete Eigenwerte. Über das Verhalten von Eigenwerten bei adiabatischen Prozessen”, *Phys. Zeit.* **30**, 467–470 (1929).
- [143] S. Ostlund, D. Rand, J. Sethna, and E. Siggia, “Universal properties of the transition from quasi-periodicity to chaos in dissipative systems”, *Physica D* **8**, 303–342 (1983).
- [144] E. Ott, “Strange attractors and chaotic motions of dynamical systems”, *Rev. Mod. Phys.* **53**, 655–671 (1981).
- [145] E. Ott, *Chaos and Dynamical Systems* (Cambridge Univ. Press, Cambridge UK, 2002).
- [146] A. Pais, *Inward Bound: of Matter and Forces in the Physical World* (Oxford Univ. Press, Oxford, 1986).
- [147] A. Pais, *Niels Bohr’s Times, in Physics, Philosophy and Polity* (Oxford Univ. Press, Oxford, 1991).
- [148] W. Parry and M. Pollicott, “An analogue of the prime number theorem for closed orbits of Axiom A flows”, *Ann. Math.* **118**, 573–591 (1983).
- [149] H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste* (Guthier-Villars, Paris, 1899).
- [150] M. Pollicott, “On the rate of mixing of Axiom A flows”, *Inv. Math.* **81**, 413–426 (1985).
- [151] M. Pollicott, “A note on the Artuso-Aurell-Cvitanović approach to the Feigenbaum tangent operator”, *J. Stat. Phys.* **62**, 257–267 (1991).
- [152] D. Rand, S. Ostlund, J. Sethna, and E. D. Siggia, “Universal transition from quasiperiodicity to chaos in dissipative systems”, *Phys. Rev. Lett.* **49**, 132–135 (1982).
- [153] V. A. Rokhlin, “Entropy of metric automorphism”, *Dokl. Akad. Nauk SSSR* **124**, 980–983 (1959).
- [154] E. Rosenqvist, *Periodic Orbit Theory Beyond Semiclassics: Convergence, Diffraction and $\hbar$ Corrections*, PhD thesis (Copenhagen Univ., Copenhagen, 1995).

- [155] D. Ruelle, “Generalized zeta-functions for Axiom A basic sets”, *Bull. Amer. Math. Soc.* **82**, 153–156 (1976).
- [156] D. Ruelle, “Zeta-functions for expanding maps and Anosov flows”, *Inv. Math.* **34**, 231–242 (1976).
- [157] D. Ruelle, “Locating resonances for Axiom A dynamical systems”, *J. Stat. Phys.* **44**, 281–292 (1986).
- [158] D. Ruelle, “One-dimensional Gibbs states and Axiom A diffeomorphisms”, *J. Diff. Geom.* **25**, 117–137 (1987).
- [159] D. Ruelle, “Resonances for Axiom A flows”, *J. Diff. Geom.* **25**, 99–116 (1987).
- [160] D. Ruelle, “The Deterministic chaos: the science and the fiction”, *Proc. R. Soc. Lond. A* **427**, 241–248 (1990).
- [161] D. Ruelle, *Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics*, 2nd ed. (Cambridge Univ. Press, Cambridge UK, 2004).
- [162] H. H. Rugh, “The correlation spectrum for hyperbolic analytic maps”, *Nonlinearity* **5**, 1237 (1992).
- [163] J. Sánchez, M. Net, B. García-Archilla, and C. Simó, “Newton-Krylov continuation of periodic orbits for Navier-Stokes flows”, *J. Comp. Phys.* **201**, 13–33 (2004).
- [164] E. Sander and J. A. Yorke, “A period-doubling cascade precedes chaos for planar maps”, *Chaos* **23**, 033113 (2013).
- [165] M. du Sautoy, *Finding Moonshine: A Mathematician’s Journey Through Symmetry* (Harper Collins, 2012).
- [166] I. Segal, “Book review: Alain Connes, Noncommutative geometry”, *Bull. Amer. Math. Soc.* **33**, 459–465 (1996).
- [167] A. Selberg, “Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series”, *J. Indian Math. Soc. (N.S.)* **20**, 47–87 (1956).
- [168] C. Shannon, “A mathematical theory of communication”, *Bell System Technical Journal* **27**, 379 (1948).
- [169] S. J. Shenker, “Scaling behavior in a map of a circle onto itself: Empirical results”, *Physica D* **5**, 405–411 (1982).
- [170] Y. G. Sinai, “On the notion of entropy of a dynamical system”, in *Selecta I*, Vol. 124 (Springer, New York, 1959), pp. 768–771.
- [171] Y. G. Sinai, “Construction of Markov partitions”, *Funct. Anal. Appl.* **2**, 245–253 (1968).
- [172] Y. G. Sinai, “Markov partitions and C-diffeomorphisms”, *Funct. Anal. Appl.* **2**, 61–82 (1968).
- [173] Y. G. Sinai, “Gibbs measures in ergodic theory”, *Russ. Math. Surv.* **27**, 21–69 (1972).

- [174] G. I. Sivashinsky, “Nonlinear analysis of hydrodynamical instability in laminar flames - I. Derivation of basic equations”, *Acta Astronaut.* **4**, 1177–1206 (1977).
- [175] J. D. Skufca, J. A. Yorke, and B. Eckhardt, “Edge of Chaos in a parallel shear flow”, *Phys. Rev. Lett.* **96**, 174101 (2006).
- [176] S. Smale, “Differentiable dynamical systems”, *Bull. Amer. Math. Soc.* **73**, 747–817 (1967).
- [177] E. A. Spiegel, “Chaos: A mixed metaphor for turbulence”, *Proc. R. Soc. Lond. A* **A413**, 87 (1987).
- [178] D. Stone, “Einstein’s unknown insight and the problem of quantizing chaos”, *Phys. Today* **58**, 37–43 (2005).
- [179] S. Toh and T. Itano, “A periodic-like solution in channel flow”, *J. Fluid Mech.* **481**, 67–76 (2003).
- [180] L. S. Tuckerman, Steady state solving via Stokes preconditioning – recursion relations for elliptic operators, in *Proc. XI Intern. Conf. Numerical Methods in Fluid Dynamics*, edited by D. L. Dwoyer, M. Y. Hussaini, and R. G. Voigt (1989), pp. 573–577.
- [181] L. van Veen, S. Kida, and G. Kawahara, “Periodic motion representing isotropic turbulence”, *Fluid Dyn. Res.* **38**, 19–46 (2006).
- [182] L. van Veen, A. Vela-Martín, and G. Kawahara, “Time-periodic inertial range dynamics”, *Phys. Rev. Lett.* **123**, 134502 (2019).
- [183] R. Vilela Mendes, “Critical-point dependence of universality in maps of the interval”, *Phys. Lett. A* **84**, 1–3 (1981).
- [184] D. Viswanath, “Recurrent motions within plane Couette turbulence”, *J. Fluid Mech.* **580**, 339–358 (2007).
- [185] E. B. Vul and K. M. Khanin, “The unstable separatrix of Feigenbaum’s fixed-point”, *Russ. Math. Surv.* **37**, 200–201 (1982).
- [186] E. B. Vul, Y. G. Sinai, and K. M. Khanin, “Feigenbaum universality and the thermodynamic formalism”, *Russ. Math. Surv.* **39**, 1–40S (1984).
- [187] F. Waleffe, “Three-dimensional coherent states in plane shear flows”, *Phys. Rev. Lett.* **81**, 4140–4148 (1998).
- [188] H. Wedin and R. R. Kerswell, “Exact coherent structures in pipe flow”, *J. Fluid Mech.* **508**, 333–371 (2004).
- [189] E. W. Weisstein, *Feigenbaum constant*, 2012.
- [190] E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1931).
- [191] D. Wintgen, K. Richter, and G. Tanner, “The semiclassical helium atom”, *Chaos* **2**, 19–33 (1992).
- [192] S. M. Zoldi and H. S. Greenside, “Spatially localized unstable periodic orbits of a high-dimensional chaotic system”, *Phys. Rev. E* **57**, R2511–R2514 (1998).

Appendix A2

Go straight

A HAMILTONIAN SYSTEM is said to be *integrable* if one can find a change of coordinates to an action-angle coordinate frame where the phase-space dynamics is described by motion on circles, one circle for each degree of freedom. In the same spirit, a natural description of a hyperbolic, unstable flow would be attained if one found a change of coordinates into a frame where the stable/unstable manifolds are straight lines, and the flow is along hyperbolas. Achieving this globally for anything but a handful of contrived examples is a pipe dream. Nevertheless, as we shall now show, we can make some headway on straightening out the flow locally.



There is much more to this story than what we touch upon here: other tricks and methods to construct regularizations, what kind of singularities could be regularized, etc.. Even though such nonlinear coordinate transformations are very important, especially in celestial mechanics, we shall not use them much in what follows, so you can safely skip this chapter on the first reading. Except, perhaps, you might like transformations that turn a Keplerian ellipse into a harmonic oscillator (example A2.2) and regularize the 2-body Coulomb collisions (sect. A2.2) in deterministic helium.

A2.1 Rectification of flows

A profitable way to exploit invariance of dynamics under smooth conjugacies is to use it to pick out the simplest possible representative of an equivalence class. These are just words, as we have no clue how to pick such ‘canonical’ representations, but for smooth flows we can always do it locally and for sufficiently short time, by appealing to the *rectification theorem*, a fundamental theorem of ordinary differential equations. The theorem tells us that a solution exists (at least for a short time interval) and what it looks like. The rectification theorem holds in the neighborhood of points of the vector field $v(x)$ that are not singular, that is, everywhere except for the equilibrium points (2.11), and points at which v is infinite.

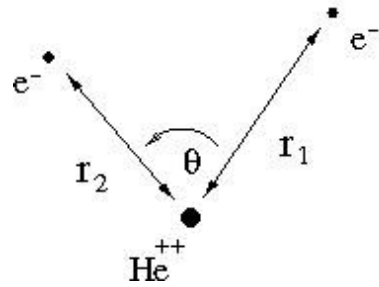


Figure A2.1: Coordinates for the helium three body problem in a plane.

According to the theorem, in a small neighborhood of a non-singular point there exists a change of coordinates $y = h(x)$ such that $\dot{x} = v(x)$ in the new, *canonical* coordinates takes form

$$\begin{aligned} \dot{y}_1 &= \dot{y}_2 = \cdots = \dot{y}_{d-1} = 0 \\ \dot{y}_d &= 1, \end{aligned} \quad (\text{A2.1})$$

with unit velocity flow along y_d , and no flow along any of the remaining directions. This is an example of a one-parameter Lie group of transformations, with the finite time τ action

$$\begin{aligned} y'_i &= y_i, & i &= 1, 2, \dots, d-1 \\ y'_d &= y_d + \tau. \end{aligned}$$

exercise 11.4
exercise A2.1



example A2.1
p. 919

A2.2 Collinear helium

(G. Tanner)

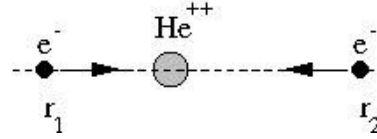


So far much has been said about 1-dimensional maps, game of pinball and other curious but rather idealized dynamical systems. If you have become impatient and started wondering what good are the methods learned so far in solving real life physical problems, good news are here. We will apply here concepts of nonlinear dynamics to nothing less than the helium, a dreaded three-body Coulomb problem.

Can we really jump from three static disks directly to three charged particles moving under the influence of their mutually attracting or repelling forces? It turns out, we can, but we have to do it with care. The full problem is indeed not accessible in all its detail, but we are able to analyze a somewhat simpler subsystem—collinear helium. This system plays an important role in the deterministic and quantum dynamics of the full three-body problem.

The deterministic helium system consists of two electrons of mass m_e and charge $-e$ moving about a positively charged nucleus of mass m_{he} and charge $+2e$.

Figure A2.2: Collinear helium, with two electrons on opposite sides of the nucleus.



The helium electron-nucleus mass ratio $m_{he}/m_e = 1836$ is so large that we may work in the infinite nucleus mass approximation $m_{he} = \infty$, fixing the nucleus at the origin. Finite nucleus mass effects can be taken into account without any substantial difficulty. We are now left with two electrons moving in three spatial dimensions around the origin. The total angular momentum of the combined electron system is still conserved. In the special case of angular momentum $L = 0$, the electrons move in a fixed plane containing the nucleus. The three body problem can then be written in terms of three independent coordinates only, the electron-nucleus distances r_1 and r_2 and the inter-electron angle Θ , see figure A2.1.

This looks like something we can lay our hands on; the problem has been reduced to three degrees of freedom, six phase-space coordinates in all, and the total energy is conserved. But let us go one step further; the electrons are attracted by the nucleus but repelled by each other. They will tend to stay as far away from each other as possible, preferably on opposite sides of the nucleus. It is thus worth having a closer look at the situation where the three particles are all on a line with the nucleus being somewhere between the two electrons. If we, in addition, let the electrons have momenta pointing towards the nucleus as in figure A2.2, then there is no force acting on the electrons perpendicular to the common interparticle axis. That is, if we start the deterministic system on the dynamical subspace $\Theta = \pi$, $\frac{d}{dt}\Theta = 0$, the three particles will remain in this *collinear configuration* for all times.

A2.2.1 Scaling

In what follows we will restrict the dynamics to this collinear subspace. It is a system of two degrees of freedom with the Hamiltonian

$$H = \frac{1}{2m_e} (p_1^2 + p_2^2) - \frac{2e^2}{r_1} - \frac{2e^2}{r_2} + \frac{e^2}{r_1 + r_2} = E, \quad (\text{A2.2})$$

where E is the total energy. As the dynamics is restricted to the fixed energy shell, the four phase-space coordinates are not independent; the energy shell dependence can be made explicit by writing

$$(r_1, r_2, p_1, p_2) \rightarrow (r_1(E), r_2(E), p_1(E), p_2(E)).$$

We will first consider the dependence of the dynamics on the energy E . A simple analysis of potential versus kinetic energy tells us that if the energy is positive both electrons can escape to $r_i \rightarrow \infty$, $i = 1, 2$. More interestingly, a single electron can still escape even if E is negative, carrying away an unlimited amount of kinetic energy, as the total energy of the remaining inner electron has no

lower bound. Not only that, but one electron *will* escape eventually for almost all starting conditions. The overall dynamics thus depends critically on whether $E > 0$ or $E < 0$. But how does the dynamics change otherwise with varying energy? Fortunately, not at all. Helium dynamics remains invariant under a change of energy up to a simple scaling transformation; a solution of the equations of motion at a fixed energy $E_0 = -1$ can be transformed into a solution at an arbitrary energy $E < 0$ by scaling the coordinates as

$$r_i(E) = \frac{e^2}{(-E)} r_i, \quad p_i(E) = \sqrt{-m_e E} p_i, \quad i = 1, 2,$$

together with a time transformation $t(E) = e^2 m_e^{1/2} (-E)^{-3/2} t$. We include the electron mass and charge in the scaling transformation in order to obtain a non-dimensionalized Hamiltonian of the form

$$H = \frac{p_1^2}{2} + \frac{p_2^2}{2} - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_1 + r_2} = -1. \quad (\text{A2.3})$$

The case of negative energies chosen here is the most interesting one for us. It exhibits chaos, unstable periodic orbits and is responsible for the bound states and resonances of the quantum problem.

A2.2.2 Regularization of two-body collisions

Next, we have a closer look at the singularities in the Hamiltonian (A2.3). Whenever two bodies come close to each other, accelerations become large, numerical routines require lots of small steps, and numerical precision suffers. No numerical routine will get us through the singularity itself, and in collinear helium electrons have no option but to collide with the nucleus. Hence a *regularization* of the differential equations of motions is a necessary prerequisite to any numerical work on such problems, both in celestial mechanics (where a spaceship executes close approaches both at the start and its destination) and in quantum mechanics (where much of semiclassical physics is dominated by returning deterministic orbits that probe the quantum wave function at the nucleus).

There is a fundamental difference between two-body collisions $r_1 = 0$ or $r_2 = 0$, and the triple collision $r_1 = r_2 = 0$. Two-body collisions can be regularized, with the singularities in equations of motion removed by a suitable coordinate transformation together with a time transformation preserving the Hamiltonian structure of the equations. Such regularization is not possible for the triple collision, and solutions of the differential equations can not be continued through the singularity at the origin. As we shall see, the chaos in collinear helium originates from this singularity of triple collisions.

A regularization of the two-body collisions is achieved by means of the Kustaanheimo–Stiefel (KS) transformation, which consists of a coordinate dependent time transformation which stretches the time scale near the origin, and a canonical transformation of the phase-space coordinates. In order to motivate the method,

we apply it first to the 1-dimensional Kepler problem

$$H = \frac{1}{2}p^2 - \frac{2}{x} = E. \quad (\text{A2.4})$$



example **A2.2**
p. **919**

We now apply this method to collinear helium. The basic idea is that one seeks a higher-dimensional generalization of the ‘square root removal’ trick (**A2.18**), by introducing a new vector Q with property $r = |Q|^2$. In this simple 1-dimensional example the KS transformation can be implemented by

$$r_1 = Q_1^2, \quad r_2 = Q_2^2, \quad p_1 = \frac{P_1}{2Q_1}, \quad p_2 = \frac{P_2}{2Q_2} \quad (\text{A2.5})$$

and reparameterization of time by $d\tau = dt/r_1 r_2$. The singular behavior in the original momenta at r_1 or $r_2 = 0$ is again compensated by stretching the time scale at these points. The Hamiltonian structure of the equations of motions with respect to the new time τ is conserved, if we consider the Hamiltonian

$$H_{ko} = \frac{1}{8}(Q_2^2 P_1^2 + Q_1^2 P_2^2) - 2R_{12}^2 + Q_1^2 Q_2^2 (-E + 1/R_{12}^2) = 0 \quad (\text{A2.6})$$

with $R_{12} = (Q_1^2 + Q_2^2)^{1/2}$, and we will take $E = -1$ in what follows. The equations of motion now have the form

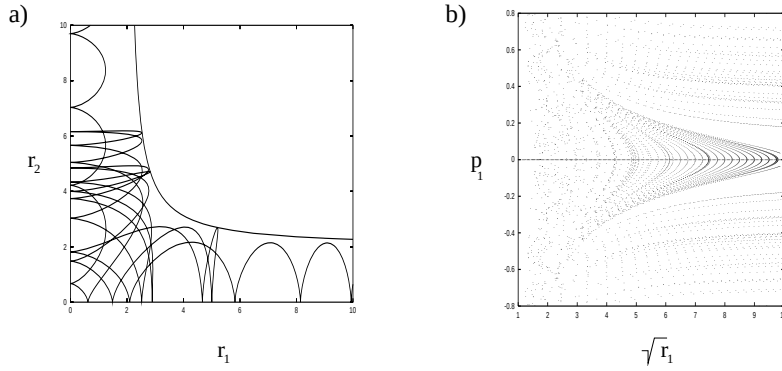
$$\begin{aligned} \dot{P}_1 &= 2Q_1 \left[2 - \frac{P_2^2}{8} - Q_2^2 \left(1 + \frac{Q_2^2}{R_{12}^4} \right) \right]; & \dot{Q}_1 &= \frac{1}{4} P_1 Q_2^2 \\ \dot{P}_2 &= 2Q_2 \left[2 - \frac{P_1^2}{8} - Q_1^2 \left(1 + \frac{Q_1^2}{R_{12}^4} \right) \right]; & \dot{Q}_2 &= \frac{1}{4} P_2 Q_1^2. \end{aligned} \quad (\text{A2.7})$$

Individual electron–nucleus collisions at $r_1 = Q_1^2 = 0$ or $r_2 = Q_2^2 = 0$ no longer pose a problem to a numerical integration routine. The equations (**A2.7**) are singular only at the triple collision $R_{12} = 0$, i.e., when both electrons hit the nucleus at the same time.

The new coordinates and the Hamiltonian (**A2.6**) are very useful when calculating trajectories for collinear helium; they are, however, less intuitive as a visualization of the three-body dynamics. We will therefore refer to the old coordinates r_1, r_2 when discussing the dynamics and the periodic orbits.

To summarize, we have brought a 3-body problem into a form where the 2-body collisions have been transformed away, and the phase-space trajectories computable numerically. To appreciate the full beauty of what has been attained, you have to study the quantum chaos part of ChaosBook.org; chapter **42**; we are already ‘almost’ ready to quantize helium by semiclassical methods.

Figure A2.3: (a) A typical trajectory in the $[r_1, r_2]$ plane; the trajectory enters here along the r_1 axis and escapes to infinity along the r_2 axis; (b) return map ($r_2=0$) for collinear helium. Strong chaos prevails for small r_1 near the nucleus.



A2.3 Rectification of maps

In sect. A2.1 we argued that nonlinear coordinate transformations can be profitably employed to simplify the representation of a flow. We shall now apply the same idea to nonlinear maps, and determine a smooth nonlinear change of coordinates that flattens out the vicinity of a fixed point and makes the map *linear* in an open neighborhood. In its simplest form the idea can be implemented only for an isolated nondegenerate fixed point (otherwise one needs the normal form expansion around the point), and only in a finite neighborhood of a point, as the conjugating function in general has a finite radius of convergence. In sect. A2.4 we will extend the method to periodic orbits.



A2.3.1 Rectification of a fixed point in one dimension

exercise A2.2

Consider a 1-dimensional map $x_{n+1} = f(x_n)$ with a fixed point at $x = 0$, with stability $\Lambda = f'(0)$. If $|\Lambda| \neq 1$, one can determine the power series for a smooth conjugation $h(x)$ centered at the fixed point, $h(0) = 0$, that flattens out the neighborhood of the fixed point

$$f(x) = h^{-1}(\Lambda h(x)) \quad (\text{A2.8})$$

and replaces the nonlinear map $f(x)$ by a *linear* map $y_{n+1} = \Lambda y_n$.

To compute the conjugation h we use the functional equation $h^{-1}(\Lambda x) = f(h^{-1}(x))$ and the expansions

$$\begin{aligned} f(x) &= \Lambda x + x^2 f_2 + x^3 f_3 + \dots \\ h^{-1}(x) &= x + x^2 h_2 + x^3 h_3 + \dots \end{aligned} \quad (\text{A2.9})$$

Equating the coefficients of x^k on both sides of the functional equation yields h_k order by order as a function of $f_2, f_3, \dots$. If $h(x)$ is a conjugation, so is any scaling $h(bx)$ of the function for a real number b . Hence the value of $h'(0)$ is not determined by the functional equation (A2.8); it is convenient to set $h'(0) = 1$.

The algebra is not particularly illuminating and best left to computers. In any case, for the time being we will not use much beyond the first, linear term in these expansions.

Here we have assumed $|\Lambda| \neq 1$. If the fixed point has vanishing $k-1$ derivatives, the conjugacy is to the k th *normal form*.

In multiple dimensions, Λ is replaced by the Jacobian matrix, and one has to check that the eigenvalues M are non-resonant, that is, there is no integer linear relation between the Floquet exponents (5.5).

remark A2.3

A2.4 Rectification of a periodic orbit

In sect. A2.3.1 we have constructed the conjugation function for a fixed point. Here we turn to the problem of constructing it for periodic orbits. Each point around the cycle has a differently distorted neighborhood, with differing second and higher order derivatives, so we need to compute a different conjugation function h_a at each periodic point x_a . We expand the map f around each periodic point along the cycle,



$$y_a(\phi) = f_a(\phi) - x_{a+1} = \phi f_{a,1} + \phi^2 f_{a,2} + \dots \quad (\text{A2.10})$$

where x_a is a point on the cycle, $f_a(\phi) = f(x_a + \phi)$ is centered on the periodic orbit, and the index k in $f_{a,k}$ refers to the k th order in the expansion (A2.9).

For a periodic orbit the conjugation formula (A2.8) generalizes to

$$f_a(\phi) = h_{a+1}^{-1}(f'_a(0)h_a(\phi)), \quad a = 1, 2, \dots, n,$$

point by point. The conjugation functions h_a are obtained in the same way as before, by equating coefficients of the expansion (A2.9), and assuming that the cycle Floquet multiplier $\Lambda = \prod_{a=0}^{n-1} f'(x_a)$ is not marginal, $|\Lambda| \neq 1$. The explicit expressions for h_a in terms of f are obtained by iterating around the whole cycle,

$$f^n(x_a + \phi) = h_a^{-1}(\Lambda h_a(\phi)) + x_a. \quad (\text{A2.11})$$

evaluated at each periodic point a . Again we have the freedom to set $h'_a(0) = 1$ for all a .

remark A2.2

A2.4.1 Repeats of cycles

We have traded our initial nonlinear map f for a (locally) linear map Λy and an equally complicated conjugation function h . What is gained by rewriting the map f in terms of the conjugacy function h ? Once the neighborhood of a fixed point

is linearized, the iterates of f are trivialized; from the conjugation formula (A2.9) one can compute the derivatives of a function composed with itself r times:

$$f^r(x) = h^{-1}(\Lambda^r h(x)).$$

One can already discern the form of the expansion for an arbitrary iterate; the answer will depend on the conjugacy function $h(x)$ computed for a *single* application of mapping f , and all the dependence on the iterate number will be carried by factors that are polynomial functions of Λ^r , a considerable simplification. The beauty of the idea is difficult to gauge at this stage—an appreciation only sets in when one starts computing perturbative corrections, whether in celestial mechanics (where the method was born), or quantum or stochastic corrections to ‘semiclassical’ approximations.



in depth:
appendix A8.1, p. 947

Résumé

The dynamical system $(\mathcal{M}, f)$ is invariant under the group of all smooth conjugacies

$$(\mathcal{M}, f) \rightarrow (\mathcal{M}', g) = (h(\mathcal{M}), h \circ f \circ h^{-1}).$$

This invariance can be used to (i) find a simplified representation for the flow and (ii) identify a set of invariants, numbers computed within a particular choice of $(\mathcal{M}, f)$, but invariant under all $\mathcal{M} \rightarrow h(\mathcal{M})$ smooth conjugacies.

The $2D$ -dimensional phase space of an integrable Hamiltonian system of D degrees of freedom is fully stratified by D -tori. In the same spirit, for a uniformly hyperbolic, chaotic dynamical system, one would like to transform to a coordinate frame in which the stable and unstable manifolds form a set of transversally intersecting hyper-planes, with the flow everywhere locally hyperbolic. That cannot be achieved in general: Fully globally integrable and fully globally chaotic flows are a very small subset of all possible flows, a ‘set of measure zero’ in the world of all dynamical systems.

What we *really* care about is developing invariant notions for a given dynamical system. The totality of smooth one-to-one nonlinear coordinate transformations h that map all trajectories of a given dynamical system $(\mathcal{M}, f^t)$ onto all trajectories of dynamical systems $(\mathcal{M}', g^t)$ gives us a huge equivalence class, much larger than the equivalence classes familiar from the theory of linear transformations. In the theory of Lie groups, the full invariant specification of an object is given by a finite set of Casimir invariants. What a good full set of invariants for a group of general nonlinear smooth conjugacies might be is not known, but the set of all periodic orbits and their Floquet multipliers turns out to be a good start.

Commentary

Remark A2.1. Rectification of flows. Consult Bluman and Kumei [4] Section 2.2.5 for a pedagogical introduction to smooth coordinate reparameterizations. Explicit examples of transformations into canonical coordinates for a group of scalings and a group of rotations are worked out.

Remark A2.2. Rectification of maps. The methods outlined above are standard in the analysis of fixed points and the construction of normal forms for bifurcations, see for example refs. [3, 5, 6, 8–10, 14–16]. The geometry underlying such methods is elegant, and we enjoyed reading, for example, Percival and Richards [13], chaps. 2 and 4 of Ozorio de Almeida’s monograph [11], and, as always, Arnol’d [2].

Recursive formulas for the evaluation of derivatives needed to evaluate (A2.9) are given, for example, in Appendix A of ref. [7]. Section 10.6 of ref. [12] describes in detail the smooth conjugacy that relates the Ulam map (14.22) to the tent map (14.21). For ‘negative Schwartzian derivatives,’ families of conjugacies of Ulam-type maps, associated Lyapunov exponents, continuous measures and further pointers to literature, see ref. [1].

Question A2.1. Is there an interpretation of different dimensions?

Q Any 3rd order equation can be written as three first order equations, the minimum needed to get chaos. Conversely, the Rössler flow can be written as a 3rd order, or *jerk equation* (do it for the Rössler equation to see what $\ddot{x}$ looks like). The axes will be acceleration, velocity, and position. My question is, which axis is acceleration, velocity and position, or could these be interchangeable (since I could have solved for $\ddot{y}$ and had the same jerk equation)?

A Oh, our first formal introduction to jerks and jounces. In dynamical systems we always express higher order derivatives as towers of first order ones, in order to be able to formulate the problem at hand as a state space + time-independent vector field at each state space point. Usually there is no interpretation of different coordinates. The passage from n th derivative to n ODEs is just a reformulation of a given problem as a dynamical systems problem, and it might be followed by any number of further linear and/or non-linear coordinate transformations, with individual coordinates retaining little or no initial physical interpretation. An example are the ‘*time delay embeddings*’, where one trades in (infinitesimal) derivatives for finite-time delays.

To be concrete, consider ChaosBook chapter 30 *Turbulence?*, example 30.6 *Equilibria of equilibria*. There we rewrite a 3rd order “jerk” ODE as a set of 3 coupled 1st order ODEs. As the original PDE equation is a 1-dimensional relative of Navier-Stokes, we make some attempts to give a physical interpretation to the new coordinates, but not very convincingly. indexjounce

Remark A2.3. A resonance condition. In the hyperbolic case there is a resonance condition that must be satisfied: none of the Floquet exponents may be related by ratios of integers. That is, if $\Lambda_{p,1}, \Lambda_{p,2}, \dots, \Lambda_{p,d}$ are the Floquet multipliers of the Jacobian matrix, then they are in resonance if there exist integers $n_1, \dots, n_d$ such that

$$(\Lambda_{p,1})^{n_1} (\Lambda_{p,2})^{n_2} \cdots (\Lambda_{p,d})^{n_d} = 1.$$

If there is resonance, one may get corrections to the basic conjugation formulas in the form of monomials in the variables of the map. (R. Mainieri)

References

- [1] J. M. Aguirregabiria, “Robust chaos with variable Lyapunov exponent in smooth one-dimensional maps”, *Chaos Solit. Fract.* **42**, 2531–2539 (2009).
- [2] V. I. Arnol’d, *Ordinary Differential Equations* (Springer, New York, 1992).
- [3] R. Belitskii, “Equivalence and normal forms of germs of smooth mappings”, *Russ. Math. Surv.* **31**, 107 (1978).
- [4] G. W. Bluman and S. Kumei, *Symmetries and Differential Equations* (Springer, New York, 1989).
- [5] A. Brjuno, “The analytic form of differential equations”, *Trans. Moscow Math. Soc.* **25**, 131–288 (1971).
- [6] T. Chen, “Equivalence and decomposition of vector fields about an elementary critical point”, *Amer. J. Math.* **85**, 10.2307/2373115 (1963).
- [7] P. Cvitanović, C. P. Dettmann, R. Mainieri, and G. Vattay, “Trace formulas for stochastic evolution operators: Weak noise perturbation theory”, *J. Stat. Phys.* **93**, 981–999 (1998).
- [8] I. Gumowski and C. Mira, *Recurrances and Discrete Dynamical Systems* (Springer, New York, 1980).
- [9] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems* (Cambridge Univ. Press, Cambridge UK, 1995).
- [10] J. Moser, “A rapidly converging iteration method and nonlinear partial differential equations I.” *Ann. Scuola Norm. Super. Pisa* **20**, 265–315 (1966).
- [11] A. M. Ozorio de Almeida, *Hamiltonian Systems: Chaos and Quantization* (Cambridge Univ. Press, Cambridge UK, 1989).
- [12] H.-O. Peitgen, H. Jürgens, and D. Saupe, *Chaos and Fractals: New Frontiers of Science* (Springer, New York, 2004).
- [13] I. Percival and D. Richards, *Introduction to Dynamics* (Cambridge Univ. Press, Cambridge, 1982).
- [14] C. L. Siegel, “Iteration of analytic functions”, *Ann. Math.* **43**, 607–612 (1942).
- [15] C. Simó, “On the analytical and numerical approximation of invariant manifolds”, in *Les Méthodes Modernes de la Mécanique Céleste (Goutelas ’89)*, edited by D. Benest and C. Froeschlé (Editions Frontières, Gif-sur-Yvette, France, 1989), pp. 285–329.
- [16] S. Sternberg, “Local contractions and a theorem of Poincaré”, *Amer. J. Math.* **79**, 809–824 (1957).

A2.5 Examples

Example A2.1. Harmonic oscillator, rectified: As a simple example of global rectification of a flow consider the harmonic oscillator

$$\dot{q} = p, \quad \dot{p} = -q. \quad (\text{A2.12})$$

The trajectories $x(t) = (q(t), p(t))$ circle around the origin, so a fair guess is that the system would have a simpler representation in polar coordinates $y = (r, \theta)$:

$$h^{-1} : \begin{cases} q &= h_1^{-1}(r, \theta) = r \cos \theta \\ p &= h_2^{-1}(r, \theta) = r \sin \theta \end{cases}. \quad (\text{A2.13})$$

The Jacobian matrix, $\partial h_i / \partial x_j$, of the transformation is

$$h' = \begin{pmatrix} \cos \theta & \sin \theta \\ -\frac{\sin \theta}{r} & \frac{\cos \theta}{r} \end{pmatrix} \quad (\text{A2.14})$$

resulting in (2.18) of rectified form

$$\begin{pmatrix} \dot{r} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\frac{\sin \theta}{r} & \frac{\cos \theta}{r} \end{pmatrix} \begin{pmatrix} \dot{q} \\ \dot{p} \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}. \quad (\text{A2.15})$$

exercise 5.1

In the new coordinates the radial coordinate r is constant, and the angular coordinate θ wraps around a cylinder with constant angular velocity. There is a subtle point in this change of coordinates: the domain of the map h^{-1} is not the plane $\mathbb{R}^2$, but rather the plane minus the origin. We mapped a plane into a cylinder, and coordinate transformations should not change the topology of the space in which the dynamics takes place; the coordinate transformation is not defined on the equilibrium point $x = (0, 0)$, or $r = 0$.

return to p. 910

Example A2.2. Keplerian ellipse, rectified: To warm up, consider the $E = 0$ case, starting at $x = 0$ at $t = 0$. Even though the equations of motion are singular at the initial point, we can immediately integrate

$$\frac{1}{2} \dot{x}^2 - \frac{2}{x} = 0$$

by means of separation of variables

$$\sqrt{x} dx = 2 dt, \quad x = (3t)^{\frac{2}{3}}, \quad (\text{A2.16})$$

and observe that the solution is not singular. The aim of regularization is to compensate for the infinite acceleration at the origin by introducing a fictitious time, in terms of which the passage through the origin is smooth.

A time transformation $dt = f(q, p)d\tau$ for a system described by a Hamiltonian $H(q, p) = E$ leaves the Hamiltonian structure of the equations of motion unaltered, if the Hamiltonian itself is transformed into $\mathcal{H}(q, p) = f(q, p)(H(q, p) - E)$. For the 1-dimensional Coulomb problem with (A2.4) we choose the time transformation $dt = x d\tau$ which lifts the $|x| \rightarrow 0$ singularity in (A2.4) and leads to a new Hamiltonian

$$\mathcal{H} = \frac{1}{2} x p^2 - 2 - E x = 0. \quad (\text{A2.17})$$

The solution (A2.16) is now parameterized by the fictitious time $d\tau$ through a pair of equations

$$x = \tau^2, \quad t = \frac{1}{3}\tau^3.$$

The equations of motion are, however, still singular as $x \rightarrow 0$:

$$\frac{d^2x}{d\tau^2} = -\frac{1}{2x} \frac{dx}{d\tau} + xE.$$

Appearance of the square root in (A2.16) now suggests a canonical transformation of form

$$x = Q^2, \quad p = \frac{P}{2Q} \tag{A2.18}$$

which maps the Kepler problem into that of a harmonic oscillator with Hamiltonian

$$H(Q, P) = \frac{1}{8}P^2 - EQ^2 = 2, \tag{A2.19}$$

with all singularities completely removed.

[return to p. 913](#)

A2.6 Exercises

A2.1. Harmonic oscillator in polar coordinates: Given a harmonic oscillator (A2.12) that follows $\dot{p} = -q$ and $\dot{q} = p$, use (A2.14) to rewrite the system in polar coordinates (A2.13) and find equations for r and θ .

1. Show that the 1-dimensional state space of the rewritten system is the quotient space $\mathcal{M}/\text{SO}(2)$.
2. Construct a Poincaré section of the quotiented flow.

A2.2. Linearization for maps. Let $f : C \rightarrow C$ be a map from the complex numbers into themselves, analytic at the origin with a fixed point. By manipulating power series, find the first few terms of the map h that conjugates f to αz , that is,

$$f(z) = h^{-1}(\alpha h(z)).$$

There are conditions on the derivative of f at the origin to assure that the conjugation is always possible. Formulate these conditions by examining the series

(difficulty: medium)

(R. Mainieri)

A2.3. Ulam and tent maps. Show that the smooth conjugacy (2.15)

$$\begin{aligned} g(y_0) &= h \circ f \circ h^{-1}(y_0) \\ y &= h(x) = \sin^2(\pi x/2), \end{aligned}$$

conjugates the tent map $f(x) = 1 - 2|x - 1/2|$ into the Ulam map $g(y) = 4y(1 - y)$. (continued as exercise 16.1)

Appendix A4

Linear stability

Mopping up operations are the activities that engage most scientists throughout their careers.

— Thomas Kuhn, *The Structure of Scientific Revolutions*

THE SUBJECT OF LINEAR ALGEBRA generates innumerable tomes of its own, and is way beyond what we can exhaustively cover. Here we recapitulate a few essential concepts that ChaosBook relies on.

The key result of is the spectral decomposition (A4.18)

$$f(\mathbf{M}) = \sum_i f(\lambda_i) P_i,$$

where

$$P_i = \prod_{j \neq i} \frac{\mathbf{M} - \lambda_j \mathbf{1}}{\lambda_i - \lambda_j}.$$

is the projection operator (A4.17) onto i th vector subspace, one for each distinct root λ_i of a matrix $\mathbf{M}$. In our applications this matrix is typically either (5.2), the equilibrium stability matrix A , or (4.24), the periodic orbit Jacobian matrix $\hat{J}$ restricted to a Poincaré section. Once the distinct non-zero eigenvalues $\{\lambda^{(i)}\}$ are computed, the associated projection operators then afford an economical description of neighborhoods of equilibria and periodic orbits, as projection operators are polynomials in $\mathbf{M}$ which need no further diagonalizations or orthogonalizations. For each distinct eigenvalue $\lambda^{(i)}$, the columns/rows of P_i are the right/left eigenvectors $\mathbf{e}^{(j)}$, $\mathbf{e}_{(j)}$ which span the corresponding linearized subspace, and are a convenient starting seed for tracing out the global unstable/stable manifolds.

While usually not phrased in language of projection operators, the requisite linear algebra is standard. We start by collecting a few standard definitions in sect. A4.1. The reader might prefer going straight to sect. A4.2.

A4.1 Linear algebra

Vector space. A set V of elements $\mathbf{x}, \mathbf{y}, \mathbf{z}, \dots$ is called a *vector* (or *linear*) *space* over a field $\mathbb{F}$ if

- (a) *vector addition* “+” is defined in V such that V is an abelian group under addition, with identity element $\mathbf{0}$;
- (b) the set is *closed* with respect to *scalar multiplication* and vector addition

$$\begin{aligned} a(\mathbf{x} + \mathbf{y}) &= a\mathbf{x} + a\mathbf{y}, & a, b \in \mathbb{F}, & \mathbf{x}, \mathbf{y} \in V \\ (a + b)\mathbf{x} &= a\mathbf{x} + b\mathbf{x} \\ a(b\mathbf{x}) &= (ab)\mathbf{x} \\ 1\mathbf{x} &= \mathbf{x}, & 0\mathbf{x} &= \mathbf{0}. \end{aligned} \tag{A4.1}$$

Here the field $\mathbb{F}$ is either $\mathbb{R}$, the field of reals numbers, or $\mathbb{C}$, the field of complex numbers. Given a subset $V_0 \subset V$, the set of all linear combinations of elements of V_0 , or the *span* of V_0 , is also a vector space.

A basis. $\{\mathbf{e}^{(1)}, \dots, \mathbf{e}^{(d)}\}$ is any linearly independent subset of V whose span is V . The number of basis elements d is the *dimension* of the vector space V .

Standard basis consists of d vectors $\{\mathbf{e}^{(1)}, \dots, \mathbf{e}^{(d)}\}$ of form $\mathbf{e}^{(1)} = (1, 0, 0, \dots, 0)$, $\mathbf{e}^{(2)} = (0, 1, 0, \dots, 0)$, $\dots$, $\mathbf{e}^{(d)} = (0, 0, 0, \dots, 1)$. The standard basis for $[d \times d]$ matrices consists of d^2 matrices with 1 as one matrix elements, the rest zero, and similarly for d^r tensors of rank r . section A10.1

Dual space, dual basis. Under a general linear transformation $g \in GL(n, \mathbb{F})$, the row of basis vectors transforms by right multiplication as $\mathbf{e}^{(j)} = \sum_k (\mathbf{g}^{-1})^j_k \mathbf{e}^{(k)}$, and the column of x_a ’s transforms by left multiplication as $x' = \mathbf{g}x$. Under left multiplication the column (row transposed) of basis vectors $\mathbf{e}_{(k)}$ transforms as $\mathbf{e}_{(j)} = (\mathbf{g}^\dagger)_j^k \mathbf{e}_{(k)}$, where the *dual rep* $\mathbf{g}^\dagger = (\mathbf{g}^{-1})^\top$ is the transpose of the inverse of $\mathbf{g}$. This observation motivates introduction of a *dual* representation space $\bar{V}$, the space on which $GL(n, \mathbb{F})$ acts via the dual rep $\mathbf{g}^\dagger$.

Definition. If V is a vector representation space, then the *dual space* $\bar{V}$ is the set of all linear forms on V over the field $\mathbb{F}$.

If $\{\mathbf{e}^{(1)}, \dots, \mathbf{e}^{(d)}\}$ is a basis of V , then $\bar{V}$ is spanned by the *dual basis* $\{\mathbf{e}_{(1)}, \dots, \mathbf{e}_{(d)}\}$, the set of d linear forms $\mathbf{e}_{(k)}$ such that

$$\mathbf{e}_{(j)} \cdot \mathbf{e}^{(k)} = \delta_j^k,$$

where δ_j^k is the Kronecker symbol, $\delta_j^k = 1$ if $j = k$, and zero otherwise. The components of dual representation space vectors $\bar{y} \in \bar{V}$ will here be distinguished by upper indices

$$(y^1, y^2, \dots, y^n). \quad (\text{A4.2})$$

They transform under $GL(n, \mathbb{F})$ as

$$y'^a = (\mathbf{g}^\dagger)^a_b y^b. \quad (\text{A4.3})$$

For $GL(n, \mathbb{F})$ no complex conjugation is implied by the † notation; that interpretation applies only to unitary subgroups $U(n) \subset GL(n, \mathbb{C})$. In the index notation, $\mathbf{g}$ can be distinguished from $\mathbf{g}^\dagger$ by keeping track of the relative ordering of the indices,

$$(\mathbf{g})_a^b \rightarrow g_a^b, \quad (\mathbf{g}^\dagger)_a^b \rightarrow g^b_a. \quad (\text{A4.4})$$

Algebra. A set of r elements $\mathbf{t}_\alpha$ of a vector space $\mathcal{T}$ forms an algebra if, in addition to the vector addition and scalar multiplication,

- (a) the set is *closed* with respect to multiplication $\mathcal{T} \cdot \mathcal{T} \rightarrow \mathcal{T}$, so that for any two elements $\mathbf{t}_\alpha, \mathbf{t}_\beta \in \mathcal{T}$, the product $\mathbf{t}_\alpha \cdot \mathbf{t}_\beta$ also belongs to $\mathcal{T}$:

$$\mathbf{t}_\alpha \cdot \mathbf{t}_\beta = \sum_{\gamma=0}^{r-1} \tau_{\alpha\beta}^\gamma \mathbf{t}_\gamma, \quad \tau_{\alpha\beta}^\gamma \in \mathbb{C}; \quad (\text{A4.5})$$

- (b) the multiplication operation is *distributive*:

$$\begin{aligned} (\mathbf{t}_\alpha + \mathbf{t}_\beta) \cdot \mathbf{t}_\gamma &= \mathbf{t}_\alpha \cdot \mathbf{t}_\gamma + \mathbf{t}_\beta \cdot \mathbf{t}_\gamma \\ \mathbf{t}_\alpha \cdot (\mathbf{t}_\beta + \mathbf{t}_\gamma) &= \mathbf{t}_\alpha \cdot \mathbf{t}_\beta + \mathbf{t}_\alpha \cdot \mathbf{t}_\gamma. \end{aligned}$$

The set of numbers $\tau_{\alpha\beta}^\gamma$ are called the *structure constants*. They form a matrix rep of the algebra,

$$(\mathbf{t}_\alpha)_\beta^\gamma \equiv \tau_{\alpha\beta}^\gamma, \quad (\text{A4.6})$$

whose dimension is the dimension r of the algebra itself.

Depending on what further assumptions one makes on the multiplication, one obtains different types of algebras. For example, if the multiplication is associative

$$(\mathbf{t}_\alpha \cdot \mathbf{t}_\beta) \cdot \mathbf{t}_\gamma = \mathbf{t}_\alpha \cdot (\mathbf{t}_\beta \cdot \mathbf{t}_\gamma),$$

the algebra is *associative*. Typical examples of products are the *matrix product*

$$(\mathbf{t}_\alpha \cdot \mathbf{t}_\beta)_a^c = (t_\alpha)_a^b (t_\beta)_b^c, \quad \mathbf{t}_\alpha \in V \otimes \bar{V}, \quad (\text{A4.7})$$

and the *Lie product*

$$(\mathbf{t}_\alpha \cdot \mathbf{t}_\beta)_a^c = (t_\alpha)_a^b (t_\beta)_b^c - (t_\alpha)_c^b (t_\beta)_b^a, \quad \mathbf{t}_\alpha \in V \otimes \bar{V} \quad (\text{A4.8})$$

which defines a *Lie algebra*.

A4.2 Eigenvalues and eigenvectors

Eigenvalues of a $[d \times d]$ matrix $\mathbf{M}$ are the roots of its characteristic polynomial

$$\det(\mathbf{M} - \lambda \mathbf{1}) = \prod (\lambda_i - \lambda) = 0. \quad (\text{A4.9})$$

Given a nonsingular matrix $\mathbf{M}$, with all $\lambda_i \neq 0$, acting on d -dimensional vectors $\mathbf{x}$, we would like to determine *eigenvectors* $\mathbf{e}^{(i)}$ of $\mathbf{M}$ on which $\mathbf{M}$ acts by scalar multiplication by eigenvalue λ_i

$$\mathbf{M} \mathbf{e}^{(i)} = \lambda_i \mathbf{e}^{(i)}. \quad (\text{A4.10})$$

If $\lambda_i \neq \lambda_j$, $\mathbf{e}^{(i)}$ and $\mathbf{e}^{(j)}$ are linearly independent. There are at most d distinct eigenvalues which we order by their real parts, $\text{Re } \lambda_i \geq \text{Re } \lambda_{i+1}$.

If all eigenvalues are distinct $\mathbf{e}^{(j)}$ are d linearly independent vectors which can be used as a (non-orthogonal) basis for any d -dimensional vector $\mathbf{x} \in \mathbb{R}^d$

$$\mathbf{x} = x_1 \mathbf{e}^{(1)} + x_2 \mathbf{e}^{(2)} + \cdots + x_d \mathbf{e}^{(d)}. \quad (\text{A4.11})$$

From (A4.10) it follows that

$$(\mathbf{M} - \lambda_i \mathbf{1}) \mathbf{e}^{(j)} = (\lambda_j - \lambda_i) \mathbf{e}^{(j)},$$

matrix $(\mathbf{M} - \lambda_i \mathbf{1})$ annihilates $\mathbf{e}^{(i)}$, thus the product of all such factors annihilates any vector, and the matrix $\mathbf{M}$ satisfies its characteristic equation (A4.9),

$$\prod_{i=1}^d (\mathbf{M} - \lambda_i \mathbf{1}) = 0. \quad (\text{A4.12})$$

This humble fact has a name: the Hamilton-Cayley theorem. If we delete one term from this product, we find that the remainder projects $\mathbf{x}$ from (A4.11) onto the corresponding eigenspace:

$$\prod_{j \neq i} (\mathbf{M} - \lambda_j \mathbf{1}) \mathbf{x} = \prod_{j \neq i} (\lambda_i - \lambda_j) x_i \mathbf{e}^{(i)}.$$

Dividing through by the $(\lambda_i - \lambda_j)$ factors yields the *projection operators*

$$P_i = \prod_{j \neq i} \frac{\mathbf{M} - \lambda_j \mathbf{1}}{\lambda_i - \lambda_j}, \quad (\text{A4.13})$$

which are *orthogonal* and *complete*:

$$P_i P_j = \delta_{ij} P_j, \quad (\text{no sum on } j), \quad \sum_{i=1}^r P_i = \mathbf{1}. \quad (\text{A4.14})$$

It follows from the characteristic equation (A4.12) that λ_i is the eigenvalue of $\mathbf{M}$ on P_i subspace:

$$\mathbf{M} P_i = \lambda_i P_i \quad (\text{no sum on } i). \quad (\text{A4.15})$$

Using $\mathbf{M} = \mathbf{M} \mathbf{1}$ and completeness relation (A4.14) we can rewrite $\mathbf{M}$ as

$$\mathbf{M} = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_d P_d. \quad (\text{A4.16})$$

with the dimension of the i th subspace given by $d_i = \text{tr } P_i$. For each distinct eigenvalue λ_i of $\mathbf{M}$,

$$(\mathbf{M} - \lambda_j \mathbf{1}) P_j = P_j (\mathbf{M} - \lambda_j \mathbf{1}) = 0, \quad (\text{A4.17})$$

the columns/rows of P_i are the right/left eigenvectors $\mathbf{e}^{(k)}$, $\mathbf{e}_{(k)}$ of $\mathbf{M}$ which (provided $\mathbf{M}$ is not of Jordan type, see example A4.2) span the corresponding linearized subspace.

The main take-home is that once the distinct non-zero eigenvalues $\{\lambda_i\}$ are computed, projection operators are polynomials in $\mathbf{M}$ which need no further diagonalizations or orthogonalizations. Any matrix function $f(\mathbf{M})$ takes the scalar value $f(\lambda_i)$ on the P_i subspace, $f(\mathbf{M}) P_i = f(\lambda_i) P_i$, and is thus easily evaluated through its *spectral decomposition*

$$f(\mathbf{M}) = \sum_i f(\lambda_i) P_i. \quad (\text{A4.18})$$

This, of course, is the reason why anyone but a fool works with irreducible reps: they reduce matrix (AKA “operator”) evaluations to manipulations with numbers.

By (A4.10) every column of P_i is proportional to a right eigenvector $\mathbf{e}^{(i)}$, and its every row to a left eigenvector $\mathbf{e}_{(i)}$. In general, neither set is orthogonal, but by the idempotence condition (A4.14), they are mutually orthogonal,

$$\mathbf{e}_{(i)} \cdot \mathbf{e}^{(j)} = c_j \delta_i^j. \quad (\text{A4.19})$$

The non-zero constant c_j is convention dependent and not worth fixing, unless you feel nostalgic about Clebsch-Gordan coefficients. We shall set $c_j = 1$. Then it is convenient to collect all left and right eigenvectors into a single matrix as follows.

Fundamental matrix (take 1). As the system is a linear, a superposition of any two solutions to $x(t) = J^t x(0)$ is also a solution. One can take any d independent initial states, $x^{(1)}(0)$, $x^{(2)}(0)$, $\dots$, $x^{(d)}(0)$, assemble them as columns of a matrix $\Phi(0)$, and formally write the solution for an arbitrary initial condition projected onto this basis,

$$x(t) = \Phi(t) \Phi(0)^{-1} x(0) \quad (\text{A4.20})$$

where $\Phi(t) = [x^{(1)}(t), x^{(2)}(t), \dots, x^{(d)}(t)]$. $\Phi(t)$ is called the *fundamental matrix* of the system, and the Jacobian matrix $J^t = \Phi(t) \Phi(0)^{-1}$ can thus be fashioned out of d trajectories $\{x^{(j)}(t)\}$. Numerically this works for sufficiently short times.

Fundamental matrix (take 2). The set of solutions $x(t) = J^t(x_0)x_0$ for a system of homogeneous linear differential equations $\dot{x}(t) = A(t)x(t)$ of order 1 and dimension d forms a d -dimensional vector space. A basis $\{\mathbf{e}^{(1)}(t), \dots, \mathbf{e}^{(d)}(t)\}$ for this vector space is called a *fundamental system*. Every solution $x(t)$ can be written as

$$x(t) = \sum_{i=1}^d c_i \mathbf{e}^{(i)}(t).$$

The $[d \times d]$ matrix $\mathbf{F}_{ij}^{-1} = \mathbf{e}_i^{(j)}$ whose columns are the right eigenvectors of J^t

$$\mathbf{F}(t)^{-1} = (\mathbf{e}^{(1)}(t), \dots, \mathbf{e}^{(d)}(t)), \quad \mathbf{F}(t)^T = (\mathbf{e}_{(1)}(t), \dots, \mathbf{e}_{(d)}(t)) \quad (\text{A4.21})$$

is the inverse of a *fundamental matrix*.

Jacobian matrix. The Jacobian matrix $J^t(x_0)$ is the linear approximation to a differentiable function $f^t(x_0)$, describing the orientation of a tangent plane to the function at a given point and the amount of local rotation and shearing caused by the transformation. The inverse of the Jacobian matrix of a function is the Jacobian matrix of the inverse function. If f is a map from d -dimensional space to itself, the Jacobian matrix is a square matrix, whose determinant we refer to as the ‘Jacobian.’

The Jacobian matrix can be written as transformation from basis at time t_0 to the basis at time t_1 ,

$$J^{t_1-t_0}(x_0) = \mathbf{F}(t_1)\mathbf{F}(t_0)^{-1}. \quad (\text{A4.22})$$

Then the matrix form of (A4.19) is $\mathbf{F}(t)\mathbf{F}(t)^{-1} = \mathbf{1}$, i.e., for zero time the Jacobian matrix is the identity.

 example A4.1
p. 933

Degenerate eigenvalues. While for a matrix with generic real elements all eigenvalues are distinct with probability 1, that is not true in presence of symmetries, or spacial parameter values (bifurcation points). What can one say about situation where d_α eigenvalues are degenerate, $\lambda_\alpha = \lambda_i = \lambda_{i+1} = \dots = \lambda_{i+d_\alpha-1}$? Hamilton-Cayley (A4.12) now takes form

$$\prod_{\alpha=1}^r (\mathbf{M} - \lambda_\alpha \mathbf{1})^{d_\alpha} = 0, \quad \sum_{\alpha} d_\alpha = d. \quad (\text{A4.23})$$

We distinguish two cases:

M can be brought to diagonal form. The characteristic equation (A4.23) can be replaced by the minimal polynomial,

$$\prod_{\alpha=1}^r (\mathbf{M} - \lambda_\alpha \mathbf{1}) = 0, \quad (\text{A4.24})$$

where the product includes each distinct eigenvalue only once. Matrix $\mathbf{M}$ acts multiplicatively

$$\mathbf{M}\mathbf{e}^{(\alpha,k)} = \lambda_i \mathbf{e}^{(\alpha,k)}, \quad (\text{A4.25})$$

on a d_α -dimensional subspace spanned by a linearly independent set of basis eigenvectors $\{\mathbf{e}^{(\alpha,1)}, \mathbf{e}^{(\alpha,2)}, \dots, \mathbf{e}^{(\alpha,d_\alpha)}\}$. This is the easy case. Luckily, if the degeneracy is due to a finite or compact symmetry group, relevant $\mathbf{M}$ matrices can always be brought to such diagonalizable form.

M can only be brought to upper-triangular, Jordan form. This is the messy case, so we only illustrate the key idea in example A4.2.



example A4.2
p. 933



example A4.3
p. 934



example A4.4
p. 935

Complex eigenvalues. As in most of our applications $\mathbf{M}$ has only real entries, it will in general have either real eigenvalues (over-damped oscillator, for example), or complex conjugate pairs of eigenvalues (under-damped oscillator, for example). The corresponding eigenvectors can be either real or complex. All coordinates used in defining the flow are real numbers, so what is the meaning of a *complex* eigenvector?

If two eigenvalues form a complex conjugate pair, $\{\lambda^{(k)}, \lambda^{(k+1)}\} = \{\mu + i\omega, \mu - i\omega\}$, they are in a sense degenerate: while a real $\lambda^{(k)}$ characterizes a motion along a line, a complex $\lambda^{(k)}$ characterizes a spiralling motion in a plane. We determine this plane by replacing the corresponding complex eigenvectors by their real and imaginary parts, $\{\mathbf{e}^{(k)}, \mathbf{e}^{(k+1)}\} \rightarrow \{\text{Re } \mathbf{e}^{(k)}, \text{Im } \mathbf{e}^{(k)}\}$, or, in terms of projection operators:

$$P_k = \frac{1}{2}(R + iQ), \quad P_{k+1} = P_k^*,$$

where $R = P_k + P_{k+1}$ is the subspace decomposed by the k th complex eigenvalue pair, and $Q = (P_k - P_{k+1})/i$, both matrices with real elements. Substitution

$$\begin{bmatrix} P_k \\ P_{k+1} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} \begin{bmatrix} R \\ Q \end{bmatrix},$$

suggest introduction of a $\det U = 1$, special unitary matrix

$$U = \frac{e^{i\pi/2}}{\sqrt{2}} \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} \quad (\text{A4.26})$$

which brings the $\lambda_k P_k + \lambda_{k+1} P_{k+1}$ complex eigenvalue pair in the spectral decomposition into the real form,

$$U^\top \begin{bmatrix} \mu + i\omega & 0 \\ 0 & \mu - i\omega \end{bmatrix} U = \begin{bmatrix} \mu & -\omega \\ \omega & \mu \end{bmatrix}.$$

$$[P_k, P_{k+1}] \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^* \end{bmatrix} \begin{bmatrix} P_k \\ P_{k+1} \end{bmatrix} = [R, Q] \begin{bmatrix} \mu & -\omega \\ \omega & \mu \end{bmatrix} \begin{bmatrix} R \\ Q \end{bmatrix}, \quad (\text{A4.27})$$

where we have dropped the superscript ^(k) for notational brevity.

exercise A4.1

To summarize, spectrally decomposed matrix **M** (A4.16) acts along lines on subspaces corresponding to real eigenvalues, and as a [2×2] rotation in a plane on subspaces corresponding to complex eigenvalue pairs.

A4.2.1 Floquet theory

When dealing with periodic orbits, some of the quantities already introduced in chapter 4 inherit names from the Floquet theory of differential equations with time-periodic coefficients. Consider the equation of variations (4.2) evaluated on a periodic orbit p of period T , at point $x(t) \in \mathcal{M}_p$,

$$\dot{\delta x} = A(t) \delta x, \quad A(t) = A(t + T),$$

with $A(t) = A(x(t))$. The periodicity of the stability matrix implies that if $\delta x(t)$ is a solution, then also $\delta x(t + T)$ satisfies the same equation: moreover the two solutions are related by (4.5)

$$\delta x(t + T) = J_p(x) \delta x(t), \quad x \in \mathcal{M}_p. \quad (\text{A4.28})$$

Even though the Jacobian matrix $J_p(x)$ depends upon x (the ‘starting’ point of the periodic orbit), we shall show in sect. 5.3 that its eigenvalues do not, so we may write the eigenvalue equation as

$$J_p(x) \mathbf{e}^{(j)}(x) = \Lambda_j \mathbf{e}^{(j)}(x), \quad (\text{A4.29})$$

where Λ_j are independent of x , and we refer to eigenvectors $\mathbf{e}^{(j)}$ as ‘covariant vectors’, or, for periodic orbits, as ‘Floquet vectors’.

Expand δx in the (A4.29) eigenbasis, $\delta x(t) = \sum \delta x_j(t) \mathbf{e}^{(j)}$, $\mathbf{e}^{(j)} = \mathbf{e}^{(j)}(x(0))$. Taking into account (A4.28), we get that $\delta x_j(t)$ is multiplied by Λ_j per each period

$$\delta x(t + T) = \sum_j \delta x_j(t + T) \mathbf{e}^{(j)} = \sum_j \Lambda_j \delta x_j(t) \mathbf{e}^{(j)}.$$

We can absorb this exponential growth / contraction by rewriting the coefficients $\delta x_j(t)$ as $\delta x_j(t) = \exp(\lambda^{(j)} t) u_j(t)$, $u_j(0) = \delta x_j(0)$. Thus each solution of the equation of variations (4.2) may be expressed in the Floquet form,

$$\delta x(t) = \sum_j e^{\lambda^{(j)} t} u_j(t) \mathbf{e}^{(j)}, \quad u_j(t + T) = u_j(t), \quad (\text{A4.30})$$

with $u_j(t)$ periodic with period T . The $\exp(\lambda^{(j)} t)$ factor is not an eigenvalue of the Jacobian matrix J^t , it is only an interpolation between x and $f^T(x)$. The continuous time t in (A4.30) does not imply that eigenvalues of the Jacobian matrix enjoy any multiplicative property for $t \neq rT$: exponents $\lambda^{(j)}$ refer to a full traversal of the periodic orbit. Indeed, while $u_j(t)$ describes the variation of $\delta x(t)$ with respect to the stationary eigen-frame fixed by eigenvectors at the point $x(0)$, the object of dynamical significance is the co-moving eigen-frame defined below in (5.11).

Table A4.1: The first 27 least stable Floquet exponents $\lambda = \mu \pm i\omega$ of equilibrium EQ_5 for plane Couette flow, $Re = 400$. Exponents are ordered by decreasing real part. The two zero exponents, to the numerical precision of our computation, arise from the two translational symmetries. For details, see ref. [5].

j	$\mu_{EQ_5}^{(j)}$	$\omega_{EQ_5}^{(j)}$	$s_1 s_2 s_3$
1,2	0.07212161	0.04074989	S S S
3	0.06209526		S A A
4	0.06162059		A S A
5,6	0.02073075	0.07355143	S S S
7	0.009925378		S A A
8,9	0.009654012	0.04551274	A A S
10,11	0.009600794	0.2302166	S A A
12,13	1.460798e-06	1.542103e-06	- - A
14,15	-0.0001343539	0.231129	A A S
16	-0.006178861		A S A
17,18	-0.007785718	0.1372092	A A S
19	-0.01064716		S A A
20,21	-0.01220116	0.2774336	S S S
22,23	-0.01539667	0.2775381	S A A
24,25	-0.03451081	0.08674062	A S A
26,27	-0.03719139	0.215319	S A A

A4.3 Eigenspectra: what to make out of them?

Well Mack the Finger said to Louie the King
 I got forty red white and blue shoe strings
 And a thousand telephones that don't ring
 Do you know where I can get rid of these things?

— [Bob Dylan](#), *Highway 61 Revisited*

Table A4.1, taken from ref. [5], is an example of how to tabulate the leading Floquet eigenvalues of the stability matrix of an equilibrium or relative equilibrium. The isotropy subgroup $G_{EQ}^{(j)}$ of the corresponding eigenfunction should be indicated. If the isotropy is trivial, $G_{EQ}^{(j)} = \{e\}$, it is omitted from the table. The isotropy subgroup G_{EQ} of the solution itself needs to be noted, and for relative equilibrium (12.19) the velocity c along the group orbit. In addition, if the least stable (i.e., the most unstable) eigenvalue is complex, it is helpful to state the period of the spiral-out motion (or spiral-in, if stable), $T_{EQ} = 2\pi/\omega_{EQ}^{(1)}$.

Table A4.2, taken from ref. [11], is an example of how to tabulate the leading Floquet exponents of the monodromy matrix of a periodic orbit or relative periodic orbit. For a periodic orbit one states the period T_p , $\Lambda_p = \prod \Lambda_{p,e}$, and the isotropy group G_p of the orbit; for a relative periodic orbit (12.23) one states in addition the shift parameters $\phi = (\phi_1, \phi_2, \dots, \phi_N)$. Λ_p , the product of expanding Floquet multipliers (5.7) is useful, as $1/|\Lambda_p|$ is the geometric weight of cycle p in a cycle expansion (remember that each complex eigenvalue contributes twice). We often do care about $\sigma_p^{(j)} = \Lambda_{p,j}/|\Lambda_{p,j}| \in \{+1, -1\}$, the sign of the j th Floquet multiplier, or, if $\Lambda_{p,j}$ is complex, its phase $T_p \omega_p^{(j)}$.

Surveying this multitude of equilibrium and Floquet exponents is aided by a

Figure A4.1: Eigenvalues of the plane Couette flow equilibrium EQ_8 , plotted according to their isotropy groups: $\bullet$ $+++$, the S -invariant subspace, $\blacktriangleright$ $+-$, $\blacktriangleleft$ $-+$, and $\blacktriangle$ $--$, where $\pm$ symbols stand for symmetric/antisymmetric under symmetry operation s_1, s_2 , and s_3 respectively, defined in ref. [4]. For tables of numerical values of stability eigenvalues see Channelflow.org.

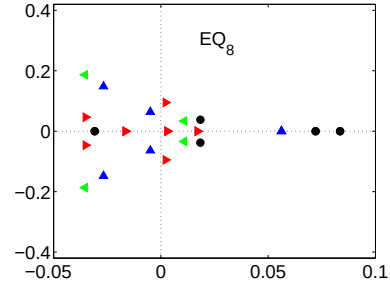


Table A4.2: The first 13 least stable Floquet exponents $\lambda = \mu \pm i\omega$ of periodic orbit $p = 59.77$ for plane Couette flow, $Re = 400$, together with the symmetries of corresponding eigenvectors. The eigenvalues are ordered by decreasing real part. The one zero eigenvalue, to the numerical precision of our computation, arises from the spanwise translational $SO(2)$ symmetry of this periodic orbit. For details, see ref. [11].

j	$\sigma_p^{(j)}$	$\mu_p^{(j)}$	$\omega_p^{(j)}$	$G_p^{(j)}$
1,2		0.07212161	0.04074989	D_1
3	1	0.06209526		?
4	-1	0.06162059		
5,6		0.02073075	0.07355143	
7	-1	0.009925378		
8,9		0.009654012	0.04551274	
10,11		0.009600794	0.2302166	

plot of the complex exponent plane (μ, ω) . An example are the eigenvalues of equilibrium EQ_8 from ref. [4], plotted in figure A4.1. To decide how many of these are “physical” in the PDE case (where number of exponents is always infinite, in principle), it is useful to look at the $(j, \mu^{(j)})$ plot. However, intelligent choice of the j -axis units can be tricky for high-dimensional problems. For Kuramoto-Sivashinsky system the correct choice are the wave-numbers which, due to the $O(2)$ symmetry, come in pairs. For plane Couette flow the good choice is not known as yet; one needs to group $O(2) \times O(2)$ wave-numbers, as well as take care of the wall-normal node counting.

Commentary

Remark A4.1. Projection operators. The construction of projection operators given in sect. A4.2 is taken from refs. [1, 2]. Sylvester [12] wrote down the spectral decomposition (A4.18) in 1883 in the form we use, but lineage certainly goes all the way back to 1795 Lagrange polynomials [10], and Euler 1783. Often projection operators get drowned in sea of algebraic details. Halmos [6] is a good early reference - but we like Harter’s exposition [7–9] best, for its multitude of specific examples and physical illustrations. In particular, by the time we get to (A4.17) we have tacitly assumed full diagonalizability of matrix $\mathbf{M}$. That is the case for the compact groups one studies in the theory of finite groups, and of compact Lie groups [3] (they are all subgroups of $U(n)$) but not necessarily in other applications. A bit of what happens then (nilpotent blocks) is touched upon in example A4.2. In his [lecture 5](#) (starts about min. 31 into the lecture) Harter explains this in great detail - its well worth your time.

References

- [1] P. Cvitanović, “Group theory for Feynman diagrams in non-Abelian gauge theories”, *Phys. Rev. D* **14**, 1536–1553 (1976).
- [2] P. Cvitanović, *Classical and exceptional Lie algebras as invariance algebras*, Oxford Univ. preprint 40/77, unpublished., 1977.
- [3] P. Cvitanović, *Group Theory: Birdtracks, Lie’s and Exceptional Groups* (Princeton Univ. Press, Princeton NJ, 2008).
- [4] J. F. Gibson, J. Halcrow, and P. Cvitanović, “Equilibrium and traveling-wave solutions of plane Couette flow”, *J. Fluid Mech.* **638**, 243–266 (2009).
- [5] J. Halcrow, *Geometry of Turbulence: An Exploration of the State-space of Plane Couette Flow*, PhD thesis (School of Physics, Georgia Inst. of Technology, Atlanta, 2008).
- [6] P. R. Halmos, *Finite-Dimensional Vector Spaces* (Princeton Univ. Press, Princeton NJ, 1948).
- [7] W. G. Harter, “Algebraic theory of ray representations of finite groups”, *J. Math. Phys.* **10**, 739–752 (1969).
- [8] W. G. Harter, *Principles of Symmetry, Dynamics, and Spectroscopy* (Wiley, New York, 1993).
- [9] W. G. Harter and N. dos Santos, “Double-group theory on the half-shell and the two-level system. I. Rotation and half-integral spin states”, *Amer. J. Phys.* **46**, 251–263 (1978).
- [10] K. Hoffman and R. Kunze, *Linear Algebra*, 2nd ed. (Prentice-Hall, Englewood Cliffs NJ, 1971).
- [11] W. Spieker, *Gone fishin’ - a blog*, School of Physics, Georgia Inst. of Technology, (Atlanta,) 2011.
- [12] J. J. Sylvester, “XXXIX. On the equation to the secular inequalities in the planetary theory”, *Philos. Mag.* **16**, 267–269 (1883).

A4.4 Examples

Example A4.1. Fundamental matrix. If A is constant in time, the system (4.2) is autonomous, and the solution is

$$x(t) = e^{At} x(0),$$

where $\exp(At)$ is defined by the Taylor series for $\exp(x)$. As the system is linear, the sum of any two solutions is also a solution. Therefore, given d independent initial conditions, $x_1(0), x_2(0), \dots, x_d(0)$ we can write the solution for an arbitrary initial condition based on its projection on to this set,

$$x(t) = \mathbf{F}(t) \mathbf{F}(0)^{-1} x(0) = e^{At} x(0),$$

where $\mathbf{F}(t) = (x_1(t), x_2(t), \dots, x_d(t))$ is a fundamental matrix of the system. (J. Halcrow)

[return to p. 927](#)

Example A4.2. Decomposition of 2-dimensional vector spaces. Enumeration of every possible kind of linear algebra eigenvalue / eigenvector combination is beyond what we can reasonably undertake here. However, enumerating solutions for the simplest case, a general $[2 \times 2]$ non-singular matrix

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

takes us a long way toward developing intuition about arbitrary finite-dimensional matrices. The eigenvalues

$$\lambda_{1,2} = \frac{1}{2} \operatorname{tr} A \pm \frac{1}{2} \sqrt{(\operatorname{tr} A)^2 - 4 \det A} \quad (\text{A4.31})$$

are the roots of the characteristic (secular) equation (A4.9):

$$\begin{aligned} \det(A - \lambda \mathbf{1}) &= (\lambda_1 - \lambda)(\lambda_2 - \lambda) \\ &= \lambda^2 - \operatorname{tr} A \lambda + \det A = 0. \end{aligned}$$

For any linear system in $\mathbb{R}^2$, there is a similarity transformation

$$B = U^{-1} A U,$$

where the columns of U consist of the generalized eigenvectors of A such that B has one of the following forms:

$$B = \begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix}, \quad B = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}, \quad B = \begin{bmatrix} \mu & -\omega \\ \omega & \mu \end{bmatrix}.$$

These three cases, called normal forms, correspond to A having (1) distinct real eigenvalues, (2) degenerate real eigenvalues, or (3) a complex pair of eigenvalues. It follows that

$$e^{Bt} = \begin{bmatrix} e^{\lambda t} & 0 \\ 0 & e^{\mu t} \end{bmatrix}, \quad e^{Bt} = e^{\lambda t} \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}, \quad e^{Bt} = e^{\mu t} \begin{bmatrix} \cos bt & -\sin bt \\ \sin bt & \cos bt \end{bmatrix},$$

where the corresponding Jacobian matrix is $e^{At} = U e^{Bt} U^{-1}$. What we have done is classify all $[2 \times 2]$ matrices as belonging to one of three classes of geometrical transformations. The first case is scaling, the second is a shear, and the third is a combination of rotation and scaling. The generalization of these normal forms to $\mathbb{R}^d$ is called the Jordan normal form.

Distinct eigenvalues case has already been described in sect. 4.8, and in the full generality for arbitrary dimension in sect. 5.1. The left/right eigenvectors are (up to overall multiplicate factors) the rows/columns of projection operators

$$P_1 = \frac{A - \lambda_2 \mathbf{1}}{\lambda_1 - \lambda_2}, \quad P_2 = \frac{A - \lambda_1 \mathbf{1}}{\lambda_2 - \lambda_1}, \quad \lambda_1 \neq \lambda_2. \quad (\text{A4.32})$$

Complex eigenvalues pair case is discussed in example 4.4

Degenerate eigenvalues. If $\lambda_1 = \lambda_2 = \lambda$, we distinguish two cases: (a) A can be brought to diagonal form. This is the easy case whose discussion in any dimension we continue in appendix A10.2.1. (b) A can be brought to *Jordan form*, with zeros everywhere except for the diagonal, and some 1's directly above it; for a $[2 \times 2]$ matrix the Jordan form is

$$A = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}, \quad \mathbf{e}^{(1)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}^{(2)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

$\mathbf{v}^{(2)}$ helps span the 2-dimensional space, $(A - \lambda)^2 \mathbf{v}^{(2)} = 0$, but is not an eigenvector, as $A \mathbf{v}^{(2)} = \lambda \mathbf{v}^{(2)} + \mathbf{e}^{(1)}$. For every such Jordan $[d_\alpha \times d_\alpha]$ block there is only one eigenvector per block. Noting that

$$A^m = \begin{bmatrix} \lambda^m & m\lambda^{m-1} \\ 0 & \lambda^m \end{bmatrix},$$

we see that instead of acting multiplicatively on $\mathbb{R}^2$, Jacobian matrix $J^t = \exp(t\mathbf{M})$

$$e^{tA} \begin{pmatrix} u \\ v \end{pmatrix} = e^{t\lambda} \begin{pmatrix} u + tv \\ v \end{pmatrix} \quad (\text{A4.33})$$

picks up a power-low correction. That spells trouble (logarithmic term $\ln t$ if we bring the extra term into the exponent). Do we care? Yes, stability matrices at bifurcation points, and for “integrable” cases such as the stadium stability (9.9) can be of this form.

(J. Halcrow)

[return to p. 928](#)

Example A4.3. Projection operator decomposition in 2 dimensions. Let's illustrate how the distinct eigenvalues case works with the $[2 \times 2]$ matrix

$$\mathbf{M} = \begin{bmatrix} 4 & 1 \\ 3 & 2 \end{bmatrix}.$$

Its eigenvalues $\{\lambda_1, \lambda_2\} = \{5, 1\}$ are the roots of (A4.31):

$$\det(\mathbf{M} - \lambda \mathbf{1}) = \lambda^2 - 6\lambda + 5 = (\lambda - 5)(\lambda - 1) = 0.$$

That $\mathbf{M}$ satisfies its secular equation (Hamilton-Cayley theorem) can be verified by explicit calculation:

$$\begin{bmatrix} 4 & 1 \\ 3 & 2 \end{bmatrix}^2 - 6 \begin{bmatrix} 4 & 1 \\ 3 & 2 \end{bmatrix} + 5 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Associated with each root λ_i is the *projection operator* (A4.32)

$$P_1 = \frac{1}{4}(\mathbf{M} - \mathbf{1}) = \frac{1}{4} \begin{bmatrix} 3 & 1 \\ 3 & 1 \end{bmatrix} \quad (\text{A4.34})$$

$$P_2 = \frac{1}{4}(\mathbf{M} - 5 \cdot \mathbf{1}) = \frac{1}{4} \begin{bmatrix} 1 & -1 \\ -3 & 3 \end{bmatrix}. \quad (\text{A4.35})$$

Matrices P_i are orthonormal and complete, The dimension of the i th subspace is given by $d_i = \text{tr } P_i$; in case at hand both subspaces are 1-dimensional. From the characteristic equation it follows that P_i satisfies the eigenvalue equation $\mathbf{M} P_i = \lambda_i P_i$. Two consequences are immediate. First, we can easily evaluate any function of $\mathbf{M}$ by spectral decomposition, for example

$$\mathbf{M}^7 - 3 \cdot \mathbf{1} = (5^7 - 3)P_1 + (1 - 3)P_2 = \begin{bmatrix} 58591 & 19531 \\ 58593 & 19529 \end{bmatrix}.$$

Second, as P_i satisfies the eigenvalue equation, its every column is a right eigenvector, and every row a left eigenvector. Picking first row/column we get the eigenvectors:

$$\begin{aligned} \{\mathbf{e}^{(1)}, \mathbf{e}^{(2)}\} &= \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \end{bmatrix} \right\} \\ \{\mathbf{e}_{(1)}, \mathbf{e}_{(2)}\} &= \left\{ \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}, \end{aligned}$$

with overall scale arbitrary. The matrix is not hermitian, so $\{\mathbf{e}^{(j)}\}$ do not form an orthogonal basis. The left-right eigenvector dot products $\mathbf{e}_{(j)} \cdot \mathbf{e}^{(k)}$, however, are orthogonal as in (A4.19), by inspection. (Continued in example 15.1.)

[return to p. 928](#)

Example A4.4. Computing matrix exponentials. If A is diagonal (the system is uncoupled), then e^{tA} is given by

$$\exp \begin{pmatrix} \lambda_1 t & & & \\ & \lambda_2 t & & \\ & & \ddots & \\ & & & \lambda_d t \end{pmatrix} = \begin{pmatrix} e^{\lambda_1 t} & & & \\ & e^{\lambda_2 t} & & \\ & & \ddots & \\ & & & e^{\lambda_d t} \end{pmatrix}.$$

If A is diagonalizable, $A = FDF^{-1}$, where D is the diagonal matrix of the eigenvalues of A and F is the matrix of corresponding eigenvectors, the result is simple: $A^n = (FDF^{-1})(FDF^{-1}) \dots (FDF^{-1}) = FD^n F^{-1}$. Inserting this into the Taylor series for e^x gives $e^{At} = F e^{Dt} F^{-1}$. But A may not have d linearly independent eigenvectors, forcing us to take a different, Jordan route, explained in example A4.2.

[return to p. 928](#)

Exercises

A4.1. **Real representation of complex eigenvalues.** (Verification of example 4.4.) λ_k, λ_{k+1} eigenvalues form a complex conjugate pair, $\{\lambda_k, \lambda_{k+1}\} = \{\mu + i\omega, \mu - i\omega\}$. Show that

- (a) corresponding projection operators are complex conjugates of each other,

$$P = P_k, \quad P^* = P_{k+1},$$

where we denote P_k by P for notational brevity.

- (b) P can be written as

$$P = \frac{1}{2}(R + iQ),$$

where $R = P_k + P_{k+1}$ and Q are matrices with real elements.

- (c)

$$\begin{bmatrix} P_k \\ P_{k+1} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} \begin{bmatrix} R \\ Q \end{bmatrix}.$$

- (d) The $\cdots + \lambda_k P_k + \lambda_k^* P_{k+1} + \cdots$ complex eigenvalue pair in the spectral decomposition (A4.16) is now replaced by a real $[2 \times 2]$ matrix

$$\cdots + \begin{bmatrix} \mu & -\omega \\ \omega & \mu \end{bmatrix} \begin{bmatrix} R \\ Q \end{bmatrix} + \cdots$$

or whatever you find the clearest way to write this real representation.

A4.2. **The matrix square root.** Consider matrix

$$A = \begin{bmatrix} 4 & 10 \\ 0 & 9 \end{bmatrix}.$$

Generalize the square root function $f(x) = x^{1/2}$ to a square root $f(A) = A^{1/2}$ of a matrix A .

- a) Which one(s) of these is/are the square root of A

$$\begin{bmatrix} 2 & 2 \\ 0 & 3 \end{bmatrix}, \begin{bmatrix} -2 & 10 \\ 0 & 3 \end{bmatrix}, \begin{bmatrix} -2 & -2 \\ 0 & -3 \end{bmatrix}, \begin{bmatrix} 2 & -10 \\ 0 & -3 \end{bmatrix}?$$

- b) Assume that the eigenvalues of a $[d \times d]$ matrix are all distinct. How many square root matrices does such matrix have?

- c) Given a $[2 \times 2]$ matrix A with a distinct pair of eigenvalues $\{\lambda_1, \lambda_2\}$, write down a formula that generates all square root matrices $A^{1/2}$. Hint: one can do this using the 2 projection operators associates with the matrix A . 2 points

Appendix A6

Lyapunov exponents

Man who says it cannot be done should not interrupt man doing it.

—Sayings of Vattay Gábor

IN THIS APPENDIX we show that the multidimensional Lyapunov exponents and relaxation exponents (dynamo rates) of vector fields can be expressed in terms of leading eigenvalues of appropriate evolution operators.

A6.1 Evolution operator for Lyapunov exponents



Lyapunov exponents were introduced and computed for 1-dimensional maps in sect. 20.4. For higher-dimensional flows only the Jacobian matrices are multiplicative, not individual eigenvalues, and the construction of the evolution operator for evaluation of the Lyapunov spectra requires the extension of evolution equations to the flow in the tangent space. We now develop the requisite theory.

Here we construct a multiplicative evolution operator (A6.4) whose spectral determinant (A6.8) yields the leading Lyapunov exponent of a d -dimensional flow (and is entire for Axiom A flows).

The key idea is to extend the dynamical system by the tangent space of the flow, suggested by the standard numerical methods for evaluation of Lyapunov exponents: start at x_0 with an initial infinitesimal tangent space vector in the d -dimensional tangent space $\eta(0) \in T\mathcal{M}_{x_0}$, and let the flow transport it along the trajectory $x(t) = f^t(x_0)$.

The dynamics in the tangent bundle $(x, \delta x) \in T\mathcal{M}$ is governed by the system of equations of variations (4.2):

$$\dot{x} = v(x), \quad \dot{\eta} = A(x)\eta.$$

Here $A(x)$ is (4.3), the stability matrix (velocity gradients matrix) of the flow. We write the solution as

$$x(t) = f^t(x_0), \quad \eta(t) = J^t(x_0) \eta_0, \quad (\text{A6.1})$$

with the tangent space vector η transported by the Jacobian matrix $J^t(x_0) = \partial x(t)/\partial x_0$ (4.5).

As explained in sect. 4.1, the growth rate of this vector is multiplicative along the trajectory and can be represented as $\eta(t) = |\eta(t)|/|\eta(0)| u(t)$ where $u(t)$ is a “unit” vector in some norm $\|\cdot\|$. For asymptotic times and for almost every initial $(x_0, \eta(0))$, this factor converges to the leading eigenvalue of the linearized stability matrix of the flow.

We implement this multiplicative evaluation of Floquet multipliers by adjoining the d -dimensional transverse tangent space $\eta \in \mathbf{T}\mathcal{M}_x$; $\eta(x) \cdot v(x) = 0$ to the $(d+1)$ -dimensional dynamical evolution space $x \in \mathcal{M} \subset \mathbb{R}^{d+1}$. In order to determine the length of the vector η we introduce a homogeneous differentiable scalar function $g(\eta) = \|\eta\|$. It has the property $g(\Lambda\eta) = |\Lambda| g(\eta)$ for any Λ . An example is the projection of a vector to its d th component

$$g \begin{pmatrix} \eta_1 \\ \eta_2 \\ \dots \\ \eta_d \end{pmatrix} = |\eta_d|.$$

Any vector $\eta(0) \in \mathbf{T}\mathcal{M}_x$ can now be represented by the product $\eta = \Lambda u$, where u is a “unit” vector in the sense that its norm is $\|u\| = 1$, and the factor

$$\Lambda^t(x_0, u_0) = g(\eta(t)) = g(J^t(x_0)u_0) \quad (\text{A6.2})$$

is the multiplicative “stretching” factor.

Unlike the leading eigenvalue of the Jacobian the stretching factor is multiplicative along the trajectory:

$$\Lambda^{t'+t}(x_0, u_0) = \Lambda^{t'}(x(t), u(t)) \Lambda^t(x_0, u_0).$$

The u evolution constrained to $E\mathbf{T}_{g,x}$, the space of unit transverse tangent vectors, is given by rescaling of (A6.1):

exercise A6.1

$$u' = R^t(x, u) = \frac{1}{\Lambda^t(x, u)} J^t(x) u. \quad (\text{A6.3})$$

Eqs. (A6.1), (A6.2) and (A6.3) enable us to define a *multiplicative* evolution operator on the extended space $U \times E\mathbf{T}_{g,x}$

$$\mathcal{L}^t(x', u'; x, u) = \delta(x' - f^t(x)) \frac{\delta(u' - R^t(x, u))}{|\Lambda^t(x, u)|^{\beta-1}}, \quad (\text{A6.4})$$

where β is a variable.

To evaluate the expectation value of $\log |\Lambda^t(x, u)|$ which is the Lyapunov exponent we again have to take the proper derivative of the leading eigenvalue of (A6.4). In order to derive the trace formula for the operator (A6.4) we need to evaluate $\text{Tr } \mathcal{L}^t = \int dx du \mathcal{L}^t(u, x; u, x)$. The $\int dx$ integral yields a weighted sum over prime periodic orbits p and their repetitions r :

$$\begin{aligned} \text{Tr } \mathcal{L}^t &= \sum_p T_p \sum_{r=1}^{\infty} \frac{\delta(t - rT_p)}{|\det(1 - M_p^r)|} \Delta_{p,r}, \\ \Delta_{p,r} &= \int_g du \frac{\delta(u - R^{T_{pr}}(x_p, u))}{|\Lambda^{T_{pr}}(x_p, u)|^{\beta-1}}, \end{aligned} \quad (\text{A6.5})$$

where M_p is the prime cycle p transverse stability matrix. As we shall see below, $\Delta_{p,r}$ is intrinsic to cycle p , and independent of any particular periodic point x_p .

We note next that if the trajectory $f^t(x)$ is periodic with period T , the tangent space contains d periodic solutions

$$\mathbf{e}^{(i)}(x(T+t)) = \mathbf{e}^{(i)}(x(t)), \quad i = 1, \dots, d,$$

corresponding to the d unit eigenvectors $\{\mathbf{e}^{(1)}, \mathbf{e}^{(2)}, \dots, \mathbf{e}^{(d)}\}$ of the transverse stability matrix, with “stretching” factors (A6.2) given by its eigenvalues

$$M_p(x)\mathbf{e}^{(i)}(x) = \Lambda_{p,i}\mathbf{e}^{(i)}(x), \quad i = 1, \dots, d. \quad (\text{no summation on } i)$$

The $\int du$ integral in (A6.5) picks up contributions from these periodic solutions. In order to compute the stability of the i th eigen-direction solution, it is convenient to expand the variation around the eigenvector $\mathbf{e}^{(i)}$ in the stability matrix eigenbasis $\delta u = \sum \delta u_\ell \mathbf{e}^{(\ell)}$. The variation of the map (A6.3) at a complete period $t = T$ is then given by

$$\begin{aligned} \delta R^T(\mathbf{e}^{(i)}) &= \frac{M\delta u}{g(M\mathbf{e}^{(i)})} - \frac{M\mathbf{e}^{(i)}}{g(M\mathbf{e}^{(i)})^2} \left(\frac{\partial g(\mathbf{e}^{(i)})}{\partial u} M\delta u \right) \\ &= \sum_{k \neq i} \frac{\Lambda_{p,k}}{\Lambda_{p,i}} \left(\mathbf{e}^{(k)} - \mathbf{e}^{(i)} \frac{\partial g(\mathbf{e}^{(i)})}{\partial u_k} \right) \delta u_k. \end{aligned} \quad (\text{A6.6})$$

The δu_i component does not contribute to this sum since $g(\mathbf{e}^{(i)} + du_i \mathbf{e}^{(i)}) = 1 + du_i$ implies $\partial g(\mathbf{e}^{(i)})/\partial u_i = 1$. Indeed, infinitesimal variations δu must satisfy

$$g(u + \delta u) = g(u) = 1 \quad \implies \quad \sum_{\ell=1}^d \delta u_\ell \frac{\partial g(u)}{\partial u_\ell} = 0,$$

so the allowed variations are of form

$$\delta u = \sum_{k \neq i} \left(\mathbf{e}^{(k)} - \mathbf{e}^{(i)} \frac{\partial g(\mathbf{e}^{(i)})}{\partial u_k} \right) c_k, \quad |c_k| \ll 1,$$

and in the neighborhood of the $\mathbf{e}^{(i)}$ eigenvector the $\int du$ integral can be expressed as

$$\int_g du = \int \prod_{k \neq i} dc_k.$$

Inserting these variations into the $\int du$ integral we obtain

$$\begin{aligned} \int_g du & \delta(\mathbf{e}^{(i)} + \delta u - R^T(\mathbf{e}^{(i)}) - \delta R^T(\mathbf{e}^{(i)}) + \dots) \\ &= \int \prod_{k \neq i} dc_k \delta((1 - \Lambda_k/\Lambda_i)c_k + \dots) \\ &= \prod_{k \neq i} \frac{1}{|1 - \Lambda_k/\Lambda_i|}, \end{aligned}$$

and the $\int du$ trace (A6.5) becomes

$$\Delta_{p,r} = \sum_{i=1}^d \frac{1}{|\Lambda_{p,i}^r|^{\beta-1}} \prod_{k \neq i} \frac{1}{|1 - \Lambda_{p,k}^r/\Lambda_{p,i}^r|}. \quad (\text{A6.7})$$

The corresponding spectral determinant is obtained by observing that the Laplace transform of the trace (21.19) is a logarithmic derivative $\text{Tr } \mathcal{L}(s) = -\frac{d}{ds} \log F(s)$ of the spectral determinant:

$$F(\beta, s) = \exp \left(- \sum_{p,r} \frac{e^{sT_p r}}{r |\det(1 - M_p^r)|} \Delta_{p,r}(\beta) \right). \quad (\text{A6.8})$$

This determinant is the central result of this section. Its zeros correspond to the eigenvalues of the evolution operator (A6.4), and can be evaluated by the cycle expansion methods.

The leading zero of (A6.8) is called ‘pressure’ (or free energy)

$$P(\beta) = s_0(\beta). \quad (\text{A6.9})$$

The average Lyapunov exponent is then given by the first derivative of the pressure at $\beta = 1$:

$$\bar{\lambda} = P'(1). \quad (\text{A6.10})$$

The simplest application of (A6.8) is to 2-dimensional hyperbolic Hamiltonian maps. The Floquet multipliers are related by $\Lambda_1 = 1/\Lambda_2 = \Lambda$, and the spectral determinant is given by

$$\begin{aligned} F(\beta, z) &= \exp \left(- \sum_{p,r} \frac{z^{rn_p}}{r |\Lambda_p^r| (1 - 1/\Lambda_p^r)^2} \Delta_{p,r}(\beta) \right) \\ \Delta_{p,r}(\beta) &= \frac{|\Lambda_p^r|^{1-\beta}}{1 - 1/\Lambda_p^{2r}} + \frac{|\Lambda_p^r|^{\beta-3}}{1 - 1/\Lambda_p^{2r}}. \end{aligned} \quad (\text{A6.11})$$

The dynamics (A6.3) can be restricted to a u unit eigenvector neighborhood corresponding to the largest eigenvalue of the Jacobi matrix. On this neighborhood the largest eigenvalue of the Jacobi matrix is the only fixed point, and the spectral determinant obtained by keeping only the largest term the $\Delta_{p,r}$ sum in (A6.7) is also entire.

In case of maps it is practical to introduce the logarithm of the leading zero and to call it ‘pressure’

$$P(\beta) = \log z_0(\beta).$$

The average of the Lyapunov exponent of the map is then given by the first derivative of the pressure at $\beta = 1$:

$$\bar{\lambda} = P'(1).$$

By factorizing the determinant (A6.11) into products of zeta functions we can conclude that the leading zero of the (A6.4) can also be recovered from the leading zeta function

$$1/\zeta_0(\beta, z) = \exp \left(- \sum_{p,r} \frac{z^{m_p}}{r|\Lambda_p^r|^\beta} \right). \quad (\text{A6.12})$$

This zeta function plays a key role in thermodynamic applications, see appendix A32.

A6.2 Advection of vector fields by chaotic flows

Fluid motions can move embedded vector fields around. An example is the magnetic field of the Sun which is “frozen” in the fluid motion. A passively evolving vector field $\mathbf{V}$ is governed by an equation of the form

$$\partial_t \mathbf{V} + \mathbf{u} \cdot \nabla \mathbf{V} - \mathbf{V} \cdot \nabla \mathbf{u} = 0, \quad (\text{A6.13})$$

where $u(x, t)$ represents the velocity field of the fluid. The strength of the vector field can grow or decay during its time evolution. The amplification of the vector field in such a process is called the “dynamo effect.” In a strongly chaotic fluid motion we can characterize the asymptotic behavior of the field with an exponent

$$\mathbf{V}(x, t) \sim \mathbf{V}(x) e^{\nu t}, \quad (\text{A6.14})$$

where ν is called the fast dynamo rate. The goal of this section is to show that periodic orbit theory can be developed for such a highly non-trivial system as well.

We can write the solution of (A6.13) formally, as shown by Cauchy. Let $\mathbf{x}(t, \mathbf{a})$ be the position of the fluid particle that was at the point $\mathbf{a}$ at $t = 0$. Then the field evolves according to

$$\mathbf{V}(\mathbf{x}, t) = \mathbf{J}(\mathbf{a}, t) \mathbf{V}(\mathbf{a}, 0) \quad , \quad (\text{A6.15})$$

where $\mathbf{J}(\mathbf{a}, t) = \partial(\mathbf{x})/\partial(\mathbf{a})$ is the Jacobian matrix of the transformation that moves the fluid into itself $\mathbf{x} = \mathbf{x}(\mathbf{a}, t)$.

We write $\mathbf{x} = f^t(\mathbf{a})$, where f^t is the flow that maps the initial positions of the fluid particles into their positions at time t . Its inverse, $\mathbf{a} = f^{-t}(\mathbf{x})$, maps particles at time t and position $\mathbf{x}$ back to their initial positions. Then we can write (A6.15)

$$V_i(\mathbf{x}, t) = \sum_j \int d^3\mathbf{a} \mathcal{L}_{ij}^t(\mathbf{x}, \mathbf{a}) V_j(\mathbf{a}, 0) \quad , \quad (\text{A6.16})$$

with

$$\mathcal{L}_{ij}^t(\mathbf{x}, \mathbf{a}) = \delta(\mathbf{a} - f^{-t}(\mathbf{x})) \frac{\partial x_i}{\partial a_j} \quad . \quad (\text{A6.17})$$

For large times, the effect of $\mathcal{L}^t$ is dominated by its leading eigenvalue, $e^{\nu_0 t}$ with $\text{Re}(\nu_0) > \text{Re}(\nu_i)$, $i = 1, 2, 3, \dots$. In this way the transfer operator furnishes the fast dynamo rate, $\nu := \nu_0$.

The trace of the transfer operator is the sum over all periodic orbit contributions, with each cycle weighted by its intrinsic stability

$$\text{Tr} \mathcal{L}^t = \sum_p T_p \sum_{r=1}^{\infty} \frac{\text{tr} M_p^r}{|\det(\mathbf{1} - M_p^{-r})|} \delta(t - rT_p). \quad (\text{A6.18})$$

We can construct the corresponding spectral determinant as usual

$$F(s) = \exp \left[- \sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{\text{tr} M_p^r}{|\det(\mathbf{1} - M_p^{-r})|} e^{srT_p} \right] \quad . \quad (\text{A6.19})$$

Note that in this formula we have omitted a term arising from the Jacobian transformation along the orbit which would give $1 + \text{tr} M_p^r$ in the numerator rather than just the trace of M_p^r . Since the extra term corresponds to advection along the orbit, and this does not evolve the magnetic field, we have chosen to ignore it. It is also interesting to note that the negative powers of the Jacobian occur in the denominator, since we have f^{-t} in (A6.17).

In order to simplify $F(s)$, we factor the denominator cycle stability determinants into products of expanding and contracting eigenvalues. For a 3-dimensional fluid flow with cycles possessing one expanding eigenvalue Λ_p (with $|\Lambda_p| > 1$), and one contracting eigenvalue λ_p (with $|\lambda_p| < 1$) the determinant may be expanded as follows:

$$|\det(\mathbf{1} - M_p^{-r})|^{-1} = |(1 - \Lambda_p^{-r})(1 - \lambda_p^{-r})|^{-1} = |\lambda_p|^r \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \Lambda_p^{-jr} \lambda_p^{kr} \quad . \quad (\text{A6.20})$$

With this decomposition we can rewrite the exponent in (A6.19) as

$$\sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{(\lambda_p^r + \Lambda_p^r) e^{srT_p}}{|\det(\mathbf{1} - M_p^{-r})|} = \sum_p \sum_{j,k=0}^{\infty} \sum_{r=1}^{\infty} \frac{1}{r} \left(|\lambda_p| \Lambda_p^{-j} \lambda_p^k e^{sT_p} \right)^r (\lambda_p^r + \Lambda_p^r) \quad , \quad (\text{A6.21})$$

which has the form of the expansion of a logarithm:

$$\sum_p \sum_{j,k} \left[\log \left(1 - e^{sT_p} |\lambda_p| \Lambda_p^{1-j} \lambda_p^k \right) + \log \left(1 - e^{sT_p} |\lambda_p| \Lambda_p^{-j} \lambda_p^{1+k} \right) \right] . \quad (\text{A6.22})$$

The spectral determinant is therefore of the form,

$$F(s) = F_e(s) F_c(s) , \quad (\text{A6.23})$$

where

$$F_e(s) = \prod_p \prod_{j,k=0}^{\infty} \left(1 - t_p^{(jk)} \Lambda_p \right) , \quad (\text{A6.24})$$

$$F_c(s) = \prod_p \prod_{j,k=0}^{\infty} \left(1 - t_p^{(jk)} \lambda_p \right) , \quad (\text{A6.25})$$

with

$$t_p^{(jk)} = e^{sT_p} |\lambda_p| \frac{\lambda_p^k}{\Lambda_p^j} . \quad (\text{A6.26})$$

The two factors present in $F(s)$ correspond to the expanding and contracting exponents. (Had we not neglected a term in (A6.19), there would be a third factor corresponding to the translation.)

For 2-dimensional Hamiltonian volume preserving systems, $\lambda = 1/\Lambda$ and (A6.24) reduces to

$$F_e(s) = \prod_p \prod_{k=0}^{\infty} \left(1 - \frac{t_p}{\Lambda_p^{k-1}} \right)^{k+1} , \quad t_p = \frac{e^{sT_p}}{|\Lambda_p|} . \quad (\text{A6.27})$$

With $\sigma_p = \Lambda_p/|\Lambda_p|$, the Hamiltonian zeta function (the $j = k = 0$ part of the product (A6.25)) is given by

$$1/\zeta_{\text{dyn}}(s) = \prod_p \left(1 - \sigma_p e^{sT_p} \right) . \quad (\text{A6.28})$$

This is a curious formula — the zeta function depends only on the return times, not on the eigenvalues of the cycles. Furthermore, the identity,

$$\frac{\Lambda + 1/\Lambda}{|(1 - \Lambda)(1 - 1/\Lambda)|} = \sigma + \frac{2}{|(1 - \Lambda)(1 - 1/\Lambda)|} ,$$

when substituted into (A6.23), leads to a relation between the vector and scalar advection spectral determinants:

$$F_{\text{dyn}}(s) = F_0^2(s)/\zeta_{\text{dyn}}(s) . \quad (\text{A6.29})$$

The spectral determinants in this equation are entire for hyperbolic (axiom A) systems, since both of them correspond to multiplicative operators.

In the case of a flow governed by a map, we can adapt the formulas (A6.27) and (A6.28) for the dynamo determinants by simply making the substitution

$$z^{n_p} = e^{sT_p} \quad , \quad (\text{A6.30})$$

where n_p is the integer order of the cycle. Then we find the spectral determinant $F_e(z)$ given by equation (A6.27) but with

$$t_p = \frac{z^{n_p}}{|\Lambda_p|} \quad (\text{A6.31})$$

for the weights, and

$$1/\zeta_{\text{dyn}}(z) = \prod_p (1 - \sigma_p z^{n_p}) \quad (\text{A6.32})$$

for the zeta-function

For *maps* with finite Markov partition the inverse zeta function (A6.32) reduces to a polynomial for z since curvature terms in the cycle expansion vanish. For example, for maps with complete binary partition, and with the fixed point stabilities of opposite signs, the cycle expansion reduces to

$$1/\zeta_{\text{dyn}}(s) = 1. \quad (\text{A6.33})$$

For such *maps* the dynamo spectral determinant is simply the square of the scalar advection spectral determinant, and therefore all its zeros are double. In other words, for flows governed by such discrete maps, the fast dynamo rate equals the scalar advection rate.

In contrast, for 3-dimensional *flows*, the dynamo effect is distinct from the scalar advection. For example, for flows with finite symbolic dynamical grammars, (A6.29) implies that the dynamo zeta function is a ratio of two entire determinants:

$$1/\zeta_{\text{dyn}}(s) = F_{\text{dyn}}(s)/F_0^2(s). \quad (\text{A6.34})$$

This relation implies that for *flows* the zeta function has double poles at the zeros of the scalar advection spectral determinant, with zeros of the dynamo spectral determinant no longer coinciding with the zeros of the scalar advection spectral determinant; Usually the leading zero of the dynamo spectral determinant is larger than the scalar advection rate, and the rate of decay of the magnetic field is no longer governed by the scalar advection.

exercise A6.2

Commentary

Remark A6.1. Lyapunov exponents. Sect. A6.1 is based on ref. [3].

Remark A6.2. Dynamo zeta. The dynamo zeta (A6.32) has been introduced by Aurell and Gilbert [1] and reviewed in ref. [4]. Our exposition follows ref. [2].

References

- [1] E. Aurell and A. D. Gilbert, “Fast dynamos and determinants of singular integral operators”, *Geophys. Astrophys. Fluid Dynam.* **73**, 5–32 (1993).
- [2] N. J. Balmforth, P. Cvitanović, G. R. Ierley, E. A. Spiegel, and G. Vattay, “Advection of vector fields by chaotic flows”, *Ann. New York Acad. Sci.* **706**, 148–160 (1993).
- [3] P. Cvitanović and G. Vattay, “Entire Fredholm determinants for evaluation of semi-classical and thermodynamical spectra”, *Phys. Rev. Lett.* **71**, 4138–4141 (1993).
- [4] A. D. Gilbert and S. Childress, *Stretch, Twist, Fold: the Fast Dynamo* (Springer, Berlin, 1995).

Exercises

A6.1. **Stretching factor.** Prove the multiplicative property of the stretching factor (A6.2). Why should we extend the phase space with the tangent space?

A6.2. **Dynamo rate.** Suppose that the fluid dynamics is highly dissipative and can be well approximated by the piecewise linear map

$$f(x) = \begin{cases} 1 + ax & \text{if } x < 0, \\ 1 - bx & \text{if } x > 0, \end{cases} \quad (\text{A6.35})$$

on an appropriate surface of section ($a, b > 2$). Suppose also that the return time is constant T_a for $x < 0$ and T_b for $x > 0$. Show that the dynamo zeta is

$$1/\zeta_{\text{dyn}}(s) = 1 - e^{sT_a} + e^{sT_b}. \quad (\text{A6.36})$$

Show also that the escape rate is the leading zero of

$$1/\zeta_0(s) = 1 - e^{sT_a}/a - e^{sT_b}/b. \quad (\text{A6.37})$$

Calculate the dynamo and the escape rates analytically if $b = a^2$ and $T_b = 2T_a$. Do the calculation for the case when you reverse the signs of the slopes of the map. What is the difference?

Appendix A8

Hamiltonian dynamics

THE SYMPLECTIC STRUCTURE of Hamilton's equations buys us much more than the incompressibility, or the phase space volume conservation alluded to in sect. 8.1.

A8.1 Stability of Hamiltonian flows



(M.J. Feigenbaum and P. Cvitanović)

The evolution equations for any p, q dependent quantity $Q = Q(q, p)$ are given by (19.29). In terms of the Poisson brackets, the time-evolution equation for $Q = Q(q, p)$ is given by (19.31). We now recast the symplectic condition (8.11) in a form convenient for using the symplectic constraints on M . Writing $x(t) = x' = [p', q']$ and the Jacobian matrix and its inverse

$$M = \begin{bmatrix} \frac{\partial q'}{\partial q} & \frac{\partial q'}{\partial p} \\ \frac{\partial p'}{\partial q} & \frac{\partial p'}{\partial p} \end{bmatrix}, \quad M^{-1} = \begin{bmatrix} \frac{\partial q}{\partial q'} & \frac{\partial q}{\partial p'} \\ \frac{\partial p}{\partial q'} & \frac{\partial p}{\partial p'} \end{bmatrix}, \quad (\text{A8.1})$$

we can spell out the symplectic invariance condition (8.11):

$$\begin{aligned} \frac{\partial q'_k}{\partial q_i} \frac{\partial p'_k}{\partial q_j} - \frac{\partial p'_k}{\partial q_i} \frac{\partial q'_k}{\partial q_j} &= 0 \\ \frac{\partial q'_k}{\partial p_i} \frac{\partial p'_k}{\partial p_j} - \frac{\partial p'_k}{\partial p_i} \frac{\partial q'_k}{\partial p_j} &= 0 \\ \frac{\partial q'_k}{\partial q_i} \frac{\partial p'_k}{\partial p_j} - \frac{\partial p'_k}{\partial q_i} \frac{\partial q'_k}{\partial p_j} &= \delta_{ij}. \end{aligned} \quad (\text{A8.2})$$

From (8.25) we obtain

$$\frac{\partial q_i}{\partial q'_j} = \frac{\partial p'_j}{\partial p_i}, \quad \frac{\partial p_i}{\partial p'_j} = \frac{\partial q'_j}{\partial q_i}, \quad \frac{\partial q_i}{\partial p'_j} = -\frac{\partial q'_j}{\partial p_i}, \quad \frac{\partial p_i}{\partial q'_j} = -\frac{\partial p'_j}{\partial q_i}. \quad (\text{A8.3})$$

Taken together, (A8.3) and (A8.2) imply that the flow conserves the $\{p, q\}$ Poisson brackets

$$\begin{aligned}\{q_i, q_j\} &= \frac{\partial q_i}{\partial p'_k} \frac{\partial q_j}{\partial q'_k} - \frac{\partial q_j}{\partial p'_k} \frac{\partial q_i}{\partial q'_k} = 0 \\ \{p_i, p_j\} &= 0, \quad \{p_i, q_j\} = \delta_{ij},\end{aligned}\tag{A8.4}$$

i.e., the transformations induced by a Hamiltonian flow are *canonical*, preserving the form of the equations of motion. The first two relations are symmetric under i, j interchange and yield $D(D-1)/2$ constraints each; the last relation yields D^2 constraints. Hence only $(2D)^2 - 2D(D-1)/2 - D^2 = d(2D+1)$ elements of M are linearly independent, as it behooves group elements of the symplectic group $\text{Sp}(2D)$.

We have now succeeded in making the full set of constraints explicit - as we shall see in appendix A31.2.1, this will enable us to implement dynamics in such a way that the symplectic invariance will be automatically preserved.

A8.2 Monodromy matrix for Hamiltonian flows

(G. Tanner)



It is not the Jacobian matrix J of the flow (4.5), but the *monodromy matrix* M , which enters the trace formula. This matrix gives the time dependence of a displacement perpendicular to the flow on the energy manifold. Indeed, we discover some trivial parts in the Jacobian matrix J . An initial displacement in the direction of the flow $x = \omega \nabla H(x)$ transfers according to $\delta x(t) = x_t(t) \delta t$ with δt time independent. The projection of any displacement on δx on $\nabla H(x)$ is constant, i.e., $\nabla H(x(t)) \delta x(t) = \delta E$. We get the equations of motion for the monodromy matrix directly choosing a suitable local coordinate system on the orbit $x(t)$ in form of the (non singular) transformation $\mathbf{U}(x(t))$:

$$\tilde{J}(x(t)) = \mathbf{U}^{-1}(x(t)) J(x(t)) \mathbf{U}(x(0))\tag{A8.5}$$

These lead to

$$\begin{aligned}\dot{\tilde{J}} &= \tilde{\mathbf{L}} \tilde{J} \\ \text{with } \tilde{\mathbf{L}} &= \mathbf{U}^{-1}(\mathbf{L}\mathbf{U} - \dot{\mathbf{U}})\end{aligned}\tag{A8.6}$$

Note that the properties a) – c) are only fulfilled for $\tilde{J}$ and $\tilde{\mathbf{L}}$ if $\mathbf{U}$ itself is symplectic.

Choosing $x_E = \nabla H(t)/|\nabla H(t)|^2$ and x_t as local coordinates uncovers the two trivial eigenvalues 1 of the transformed matrix in (A8.5) at any time t . Setting

$\mathbf{U} = (x_t^\top, x_E^\top, x_1^\top, \dots, x_{2d-2}^\top)$ gives

$$\tilde{\mathbf{J}} = \begin{pmatrix} 1 & * & * & \dots & * \\ 0 & 1 & 0 & \dots & 0 \\ 0 & * & & & \\ \vdots & \vdots & & \mathbf{M} & \\ 0 & * & & & \end{pmatrix}; \quad \tilde{\mathbf{L}} = \begin{pmatrix} 0 & * & * & \dots & * \\ 0 & 0 & 0 & \dots & 0 \\ 0 & * & & & \\ \vdots & \vdots & & \mathbf{I} & \\ 0 & * & & & \end{pmatrix}, \quad (\text{A8.7})$$

The matrix $\mathbf{M}$ is now the monodromy matrix and the equation of motion are given by

$$\dot{\mathbf{M}} = \mathbf{I} \mathbf{M}. \quad (\text{A8.8})$$

The vectors $x_1, \dots, x_{2d-2}$ must span the space perpendicular to the flow on the energy manifold.

For a system with two degrees of freedom, the matrix $\mathbf{U}(\mathbf{t})$ can be written down explicitly, i.e.,

$$\mathbf{U}(t) = (x_t, x_1, x_E, x_2) = \begin{pmatrix} \dot{x} & -\dot{y} & -\dot{u}/q^2 & -\dot{v}/q^2 \\ \dot{y} & \dot{x} & -\dot{v}/q^2 & \dot{u}/q^2 \\ \dot{u} & \dot{v} & \dot{x}/q^2 & -\dot{y}/q^2 \\ \dot{v} & -\dot{u} & \dot{y}/q^2 & \dot{x}/q^2 \end{pmatrix} \quad (\text{A8.9})$$

with $x^\top = (x, y; u, v)$ and $q = |\nabla H| = |\dot{x}|$. The matrix $\mathbf{U}$ is non singular and symplectic at every phase space point x , except the equilibrium points $\dot{x} = 0$. The matrix elements for $\mathbf{I}$ are given (A8.11). One distinguishes 4 classes of eigenvalues of $\mathbf{M}$.

- *stable* or *elliptic*, if $\Lambda = e^{\pm i\pi\nu}$ and $\nu \in]0, 1[$.
- *marginal*, if $\Lambda = \pm 1$.
- *hyperbolic*, *inverse hyperbolic*, if $\Lambda = e^{\pm\lambda}$, $\Lambda = -e^{\pm\lambda}$.
- *loxodromic*, if $\Lambda = e^{\pm\mu \pm i\omega}$ with μ and ω real. This is the most general case, possible only in systems with 3 or more degrees of freedom.

For 2 degrees of freedom, i.e., $\mathbf{M}$ is a $[2 \times 2]$ matrix, the eigenvalues are determined by

$$\lambda = \frac{\text{tr}(\mathbf{M}) \pm \sqrt{\text{tr}(\mathbf{M})^2 - 4}}{2}, \quad (\text{A8.10})$$

i.e., $\text{tr}(\mathbf{M}) = 2$ separates stable and unstable behavior.

The $\mathbf{I}$ matrix elements for the local transformation (A8.9) are

$$\begin{aligned}
 \tilde{\mathbf{I}}_{11} &= \frac{1}{q} [(h_x^2 - h_y^2 - h_u^2 + h_v^2)(h_{xu} - h_{yv}) + 2(h_x h_y - h_u h_v)(h_{xv} + h_{yu}) \\
 &\quad - (h_x h_u + h_y h_v)(h_{xx} + h_{yy} - h_{uu} - h_{vv})] \\
 \tilde{\mathbf{I}}_{12} &= \frac{1}{q^2} [(h_x^2 + h_y^2)(h_{yy} + h_{uu}) + (h_y^2 + h_u^2)(h_{xx} + h_{vv}) \\
 &\quad - 2(h_x h_u + h_y h_v)(h_{xu} + h_{yv}) - 2(h_x h_y - h_u h_v)(h_{xy} - h_{uv})] \\
 \tilde{\mathbf{I}}_{21} &= -(h_x^2 + h_y^2)(h_{uu} + h_{vv}) - (h_u^2 + h_v^2)(h_{xx} + h_{yy}) \\
 &\quad + 2(h_x h_u - h_y h_v)(h_{xu} - h_{yv}) + 2(h_x h_v + h_y h_u)(h_{xv} + h_{yu}) \\
 \tilde{\mathbf{I}}_{22} &= -\tilde{\mathbf{I}}_{11},
 \end{aligned} \tag{A8.11}$$

with h_i, h_{ij} is the derivative of the Hamiltonian H with respect to the phase space coordinates and $q = |\nabla H|^2$.

Appendix A10

Flips, slides and turns

BASIC GROUP-THEORETIC NOTIONS are recapitulated here: groups, irreducible representations, invariants. Our notation follows birdtracks.eu.

The key result is the construction of projection operators from invariant matrices. The basic idea is simple: a hermitian matrix can be diagonalized. If this matrix is an invariant matrix, it decomposes the reps of the group into direct sums of lower-dimensional reps. Most of computations to follow implement the spectral decomposition

$$M = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_r P_r,$$

which associates with each distinct root λ_i of invariant matrix M a projection operator (A10.20):

$$P_i = \prod_{j \neq i} \frac{M - \lambda_j \mathbf{1}}{\lambda_i - \lambda_j}.$$

Sects. A24.2 and A24.3 develop Fourier analysis as an application of the general theory of invariance groups and their representations.

A10.1 Preliminaries and definitions

(A. Wirzba and P. Cvitanović)

We define *group*, *representation*, *symmetry of a dynamical system*, and *invariance*.

Group axioms. A group G is a set of elements $g_1, g_2, g_3, \dots$ for which *composition* or *group multiplication* $g_2 \circ g_1$ (which we often abbreviate as $g_2 g_1$) of any two elements satisfies the following conditions:

1. If $g_1, g_2 \in G$, then $g_2 \circ g_1 \in G$.
2. The group multiplication is associative: $g_3 \circ (g_2 \circ g_1) = (g_3 \circ g_2) \circ g_1$.
3. The group G contains *identity* element e such that $g \circ e = e \circ g = g$ for every element $g \in G$.
4. For every element $g \in G$, there exists a unique $h == g^{-1} \in G$ such that $h \circ g = g \circ h = e$.

A *finite* group is a group with a finite number of elements

$$G = \{e, g_2, \dots, g_{|G|}\},$$

where $|G|$, the number of elements, is the *order* of the group.

Groups are defined and classified as abstract objects by their multiplication tables (for finite groups) or Lie algebras (for Lie groups). What concerns us in applications is their *action* as groups of transformations on a given space, usually a vector space (see appendix A4.1), but sometimes an affine space, or a more general manifold $\mathcal{M}$.

Repeated index summation. Throughout this text, the repeated pairs of upper/lower indices are always summed over

$$G_a{}^b x_b \equiv \sum_{b=1}^n G_a{}^b x_b, \quad (\text{A10.1})$$

unless explicitly stated otherwise.

General linear transformations. Let $GL(n, \mathbb{F})$ be the group of general linear transformations,

$$GL(n, \mathbb{F}) = \{g : \mathbb{F}^n \rightarrow \mathbb{F}^n \mid \det(g) \neq 0\}. \quad (\text{A10.2})$$

Under $GL(n, \mathbb{F})$ a basis set of V is mapped into another basis set by multiplication with a $[n \times n]$ matrix g with entries in field $\mathbb{F}$ ($\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$),

$$\mathbf{e}'^a = \mathbf{e}^b (g^{-1})_b{}^a.$$

As the vector $\mathbf{x}$ is what it is, regardless of a particular choice of basis, under this transformation its coordinates must transform as

$$x'_a = g_a{}^b x_b.$$

Standard rep. We shall refer to the set of $[n \times n]$ matrices g as a *standard rep* of $GL(n, \mathbb{F})$, and the space of all n -tuples $(x_1, x_2, \dots, x_n)^\top$, $x_i \in \mathbb{F}$ on which these matrices act as the *standard representation space* V .

Under a general linear transformation $g \in GL(n, \mathbb{F})$, the row of basis vectors transforms by right multiplication as $\mathbf{e}' = \mathbf{e} g^{-1}$, and the column of x_a 's transforms by left multiplication as $x' = g x$. Under left multiplication the column (row transposed) of basis vectors $\mathbf{e}^\top$ transforms as $\mathbf{e}'^\top = g^\dagger \mathbf{e}^\top$, where the *dual rep* $g^\dagger = (g^{-1})^\top$ is the transpose of the inverse of g . This observation motivates introduction of a *dual representation space* $\bar{V}$, the space on which $GL(n, \mathbb{F})$ acts via the dual rep $g^\dagger$.

Dual space. If V is a vector representation space, then the *dual space* $\bar{V}$ is the set of all linear forms on V over the field $\mathbb{F}$.

If $\{\mathbf{e}^{(1)}, \dots, \mathbf{e}^{(d)}\}$ is a (right) basis of V , then $\bar{V}$ is spanned by the *dual basis* (left basis) $\{\mathbf{e}_{(1)}, \dots, \mathbf{e}_{(d)}\}$, the set of n linear forms $\mathbf{e}_{(j)}$ such that

$$\mathbf{e}_{(i)} \cdot \mathbf{e}^{(j)} = \delta_i^j,$$

where δ_a^b is the Kronecker symbol, $\delta_a^b = 1$ if $a = b$, and zero otherwise. The components of dual representation space vectors will here be distinguished by upper indices

$$(y^1, y^2, \dots, y^n). \quad (\text{A10.3})$$

They transform under $GL(n, \mathbb{F})$ as

$$y'^a = (g^\dagger)_b^a y^b. \quad (\text{A10.4})$$

For $GL(n, \mathbb{F})$ no complex conjugation is implied by the † notation; that interpretation applies only to unitary subgroups of $GL(n, \mathbb{C})$. g can be distinguished from $g^\dagger$ by meticulously keeping track of the relative ordering of the indices,

$$g_a^b \rightarrow g_a^b, \quad (g^\dagger)_a^b \rightarrow g^b_a. \quad (\text{A10.5})$$

Defining space, dual space. In what follows V will always denote the *defining* n -dimensional complex vector representation space, that is to say the initial, “elementary multiplet” space within which we commence our deliberations. Along with the defining vector representation space V comes the *dual* n -dimensional vector representation space $\bar{V}$. We shall denote the corresponding element of $\bar{V}$ by raising the index, as in (A10.3), so the components of defining space vectors, resp. dual vectors, are distinguished by lower, resp. upper indices:

$$\begin{aligned} x &= (x_1, x_2, \dots, x_n), & \mathbf{x} &\in V \\ \bar{x} &= (x^1, x^2, \dots, x^n), & \bar{\mathbf{x}} &\in \bar{V}. \end{aligned} \quad (\text{A10.6})$$

Defining rep. Let G be a group of transformations acting linearly on V , with the action of a group element $g \in G$ on a vector $x \in V$ given by an $[n \times n]$ matrix g

$$x'_a = g_a^b x_b \quad a, b = 1, 2, \dots, n. \quad (\text{A10.7})$$

We shall refer to g_a^b as the *defining rep* of the group G . The action of $g \in G$ on a vector $\bar{q} \in \bar{V}$ is given by the *dual rep* $[n \times n]$ matrix $g^\dagger$:

$$x'^a = x^b (g^\dagger)_b^a = g^a_b x^b. \quad (\text{A10.8})$$

In the applications considered here, the group G will almost always be assumed to be a subgroup of the *unitary group*, in which case $g^{-1} = g^\dagger$, and † indicates hermitian conjugation:

$$(g^\dagger)_a^b = (g_b^a)^* = g^b_a. \quad (\text{A10.9})$$

Hermitian conjugation is effected by complex conjugation and index transposition: Complex conjugation interchanges upper and lower indices; transposition reverses their order. A matrix is *hermitian* if its elements satisfy

$$(M^\dagger)_b^a = M_b^a. \quad (\text{A10.10})$$

For a hermitian matrix there is no need to keep track of the relative ordering of indices, as $M_b^a = (M^\dagger)_b^a = M^a_b$.

Invariant vectors. The vector $q \in V$ is an *invariant vector* if for any transformation $g \in G$

$$q = gq. \quad (\text{A10.11})$$

If a bilinear form $M(\bar{x}, y) = x^a M_a^b y_b$ is invariant for all $g \in G$, the matrix

$$M_a^b = g_a^c g^b_d M_c^d \quad (\text{A10.12})$$

is an *invariant matrix*. Multiplying with g_b^e and using the unitary condition (A10.9), we find that the invariant matrices *commute* with all transformations $g \in G$:

$$[g, M] = 0. \quad (\text{A10.13})$$

Invariants. We shall refer to an invariant relation between p vectors in V and q vectors in $\bar{V}$, which can be written as a homogeneous polynomial in terms of vector components, such as

$$H(x, y, \bar{z}, \bar{r}, \bar{s}) = h^{ab}_{cde} x_b y_a s^e r^d z^c, \quad (\text{A10.14})$$

as an *invariant* in $V^q \otimes \bar{V}^p$ (repeated indices, as always, summed over). In this example, the coefficients h^{ab}_{cde} are components of invariant tensor $h \in V^3 \otimes \bar{V}^2$.

Matrix representation of a group. Let us now map the abstract group G *homeomorphically* on a group of matrices $\mathbf{D}(G)$ acting on the vector space V , i.e., in such a way that the group properties, especially the group multiplication, are preserved:

1. Any $g \in G$ is mapped to a matrix $\mathbf{D}(g) \in \mathbf{D}(G)$.
2. The group product $g_2 \circ g_1 \in G$ is mapped onto the matrix product $\mathbf{D}(g_2 \circ g_1) = \mathbf{D}(g_2)\mathbf{D}(g_1)$.
3. The associativity follows from the associativity of matrix multiplication:
 $\mathbf{D}(g_3 \circ (g_2 \circ g_1)) = \mathbf{D}(g_3)(\mathbf{D}(g_2)\mathbf{D}(g_1)) = (\mathbf{D}(g_3)(\mathbf{D}(g_2))\mathbf{D}(g_1)$.
4. The identity element $e \in G$ is mapped onto the unit matrix $\mathbf{D}(e) = \mathbf{1}$ and the inverse element $g^{-1} \in G$ is mapped onto the inverse matrix $\mathbf{D}(g^{-1}) = [\mathbf{D}(g)]^{-1} \equiv \mathbf{D}^{-1}(g)$.

We call this matrix group $\mathbf{D}(G)$ a linear or matrix *representation* of the group G in the *representation space* V . We emphasize here ‘linear’ in order to distinguish the matrix representations from other representations that do not have to be linear, in general. Throughout this appendix we only consider linear representations.

If the dimensionality of V is d , we say the representation is an *d-dimensional representation*. We will often abbreviate the notation by writing matrices $\mathbf{D}(g) \in \mathbf{D}(G)$ as g , i.e., $x' = gx$ corresponds to the matrix operation $x'_i = \sum_{j=1}^d \mathbf{D}(g)_{ij}x_j$.

Faithful representations, factor group. If the mapping G on $\mathbf{D}(G)$ is an isomorphism, the representation is said to be *faithful*. In this case the order of the group of matrices $\mathbf{D}(G)$ is equal to the order $|G|$ of the group. In general, however, there will be several elements $h \in G$ that will be mapped on the unit matrix $\mathbf{D}(h) = \mathbf{1}$. This property can be used to define a subgroup $H \subset G$ of the group G consisting of all elements $h \in G$ that are mapped to the unit matrix of a given representation. Then the representation is a faithful representation of the *factor group* G/H .

Equivalent representations, equivalence classes. A representation of a group is by no means unique. If the basis in the d -dimensional vector space V is changed, the matrices $\mathbf{D}(g)$ have to be replaced by their transformations $\mathbf{D}'(g)$, with the new matrices $\mathbf{D}'(g)$ and the old matrices $\mathbf{D}(g)$ are related by an equivalence transformation through a non-singular matrix $\mathbf{C}$

$$\mathbf{D}'(g) = \mathbf{C}\mathbf{D}(g)\mathbf{C}^{-1}.$$

The group of matrices $\mathbf{D}'(g)$ form a representation $\mathbf{D}'(G)$ equivalent to the representation $\mathbf{D}(G)$ of the group G . The equivalent representations have the same structure, although the matrices look different.

Character of a representation. The character of $\chi_\mu(g)$ of a d -dimensional representation $\mathbf{D}(g)$ of the group element $g \in G$ is defined as trace

$$\chi_\mu(g) = \text{tr } \mathbf{D}(g) = \sum_{i=1}^d \mathbf{D}_{ii}(g).$$

Note that $\chi(e) = d$, since $\mathbf{D}_{ij}(e) = \delta_{ij}$ for $1 \leq i, j \leq d$.

Because of the cyclic nature of the trace the character of equivalent representations is the same

$$\chi(g) = \sum_{i=1}^n \mathbf{D}'_{ii}(g) = \text{tr } \mathbf{D}'(g) = \text{tr } (\mathbf{C}\mathbf{D}(g)\mathbf{C}^{-1}).$$

Definition: Character tables. Finding a transformation S which simultaneously block-diagonalizes the regular representation of each group element sounds difficult. However, suppose it can be achieved, and we obtain a set of irreps $D^{(\mu)}(g)$, then according to Schur's lemmas, $D^{(\mu)}(g)$ must satisfy a set of orthogonality relations:

$$\frac{d_\mu}{|G|} \sum_g D_{il}^{(\mu)}(g) D_{mj}^{(\nu)}(g^{-1}) = \delta_{\mu\nu} \delta_{ij} \delta_{lm}. \quad (\text{A10.15})$$

Denote the trace of irrep $D_{il}^{(\mu)}$ as $\chi(\mu)$, and we call it the character of $D^{(\mu)}$. Properties of irreps can be derived from (A10.15), and we list them as follows:

1. The number of irreps is the same as the number of classes.
2. Dimensions of irreps satisfy $\sum_{\mu=1}^r d_\mu^2 = |G|$
3. orthonormal relation I : $\sum_i |K_i| \chi_i^{(\mu)} \chi_i^{(\nu)*} = |G| \delta_{\mu\nu}$.
Here, the summation goes through all classes of this group, and $|K_i|$ is the number of elements in class i . This weight comes from the fact that elements in the same class have the same character.
4. orthonormal relation II : $\sum_\mu \chi_i^{(\mu)} \chi_j^{(\mu)*} = \frac{|G|}{|K_i|} \delta_{ij}$.

The characters for all classes and irreps of a finite group are conventionally arranged into a *character table*, a square array whose rows represent different classes and columns represent different irreps (as usual, 50% of authors, including ChaosBook, will use columns and rows instead). Rules 1 and 2 help determine the number of irreps and their dimensions. As matrix representation of class $\{e\}$ is always the identity matrix, the first column is always the dimension of the corresponding representation. All entries of the first row are always 1, because the symmetric irrep is always 1-dimensional. To compute the remaining entries, use properties 3, 4 and the class multiplication tables.

Definition: Projection operators. We have listed the properties of irreps and the techniques of constructing character table, but we still do not know how to construct the similarity transformation S which takes a regular representation into a block-diagonal form.

One of these invariant subspace is $|G|^{-1} \sum_g \rho(gx)$, which is the basis of the 1-d symmetric irrep A . For C_3 , it is (25.38). But how to get others? We need to resort to the projection operator:

$$P_i^{(\mu)} = \frac{d_\mu}{|G|} \sum_g D_{ii}^{(\mu)}(g) U(g) \quad (\text{A10.16})$$

It projects an arbitrary function into the i_{th} basis of irrep $D^{(\mu)}$ provided the diagonal elements of this representation $D_{ii}^{(\mu)}$ is known. $P_i^{(\mu)} \rho(x) = \rho_i^{(\mu)}$.

For 1-dimensional representations, this projection operator is known after we obtain the character table, since character of 1-d matrix is the matrix itself. But for 2-dimensional or higher dimensional representations, we need to know the diagonal elements $D_{ii}^{(\mu)}$ in order to get the bases of invariant subspaces.

Summing i in (A10.16) gives

$$P^{(\mu)} = \frac{d_\mu}{|G|} \sum_g \chi^{(\mu)}(g) U(g) \quad (\text{A10.17})$$

This is also a projection operator which projects an arbitrary function onto the sum of basis of irrep $D^{(\mu)}$. We will use this operator to split the trace of evolution operator into sum over all different irreps.

A10.2 Invariants and reducibility

What follows is a bit dry, so we start with a motivational quote from Hermann Weyl on the “so-called first main theorem of invariant theory”:

“All invariants are expressible in terms of a finite number among them. We cannot claim its validity for every group G ; rather, it will be our chief task to investigate for each particular group whether a finite integrity basis exists or not; the answer, to be sure, will turn out affirmative in the most important cases.”

It is easy to show that any rep of a finite group can be brought to unitary form, and the same is true of all compact Lie groups. Hence, in what follows, we specialize to unitary and hermitian matrices.

A10.2.1 Projection operators

For $\mathbf{M}$ a hermitian matrix, there exists a diagonalizing unitary matrix $\mathbf{C}$ such that

$$\mathbf{CMC}^\dagger = \begin{bmatrix} \lambda_1 & \dots & 0 & & & \\ & \ddots & & & & \\ 0 & \dots & \lambda_1 & & & \\ \hline & & & \lambda_2 & 0 & \dots & 0 \\ & & & 0 & \lambda_2 & & \\ & 0 & & \vdots & & \ddots & \\ & & & 0 & \dots & \lambda_2 & \\ \hline & & & & & & \lambda_3 & \dots \\ & 0 & & & 0 & & \vdots & \ddots \end{bmatrix}. \quad (\text{A10.18})$$

Here $\lambda_i \neq \lambda_j$ are the r distinct roots of the minimal *characteristic* (or *secular*) polynomial

$$\prod_{i=1}^r (\mathbf{M} - \lambda_i \mathbf{1}) = 0. \quad (\text{A10.19})$$

In the matrix $\mathbf{C}(\mathbf{M} - \lambda_2 \mathbf{1})\mathbf{C}^\dagger$ the eigenvalues corresponding to λ_2 are replaced by zeroes:

$$\begin{bmatrix} \lambda_1 - \lambda_2 & & & & & \\ & \lambda_1 - \lambda_2 & & & & \\ \hline & & 0 & & & \\ & & & \ddots & & \\ & & & & 0 & \\ \hline & & & & & \lambda_3 - \lambda_2 & & \\ & & & & & \lambda_3 - \lambda_2 & & \\ & & & & & & \ddots & \end{bmatrix},$$

and so on, so the product over all factors $(\mathbf{M} - \lambda_2 \mathbf{1})(\mathbf{M} - \lambda_3 \mathbf{1}) \dots$, with exception of the $(\mathbf{M} - \lambda_1 \mathbf{1})$ factor, has nonzero entries only in the subspace associated with λ_1 :

$$\mathbf{C} \prod_{j \neq 1} (\mathbf{M} - \lambda_j \mathbf{1}) \mathbf{C}^\dagger = \prod_{j \neq 1} (\lambda_1 - \lambda_j) \begin{bmatrix} 1 & 0 & 0 & & & \\ 0 & 1 & 0 & & & \\ 0 & 0 & 1 & & & \\ \hline & & & 0 & & \\ & & & & 0 & \\ & 0 & & & & 0 & \\ & & & & & & \ddots \end{bmatrix}.$$

Thus we can associate with each distinct root λ_i a *projection operator* P_i ,

$$P_i = \prod_{j \neq i} \frac{\mathbf{M} - \lambda_j \mathbf{1}}{\lambda_i - \lambda_j}, \quad (\text{A10.20})$$

which acts as identity on the i th subspace, and zero elsewhere. For example, the projection operator onto the λ_1 subspace is

$$P_1 = \mathbf{C}^\dagger \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & \text{||||} & & \\ \hline & & & 0 & \\ & & & & 0 \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix} \mathbf{C}. \quad (\text{A10.21})$$

The diagonalization matrix $\mathbf{C}$ is deployed in the above only as a pedagogical device. The whole point of the projector operator formalism is that we *never* need to carry such explicit diagonalization; all we need are whatever invariant matrices $\mathbf{M}$ we find convenient, the algebraic relations they satisfy, and orthonormality and completeness of P_i : The matrices P_i are *orthogonal*

$$P_i P_j = \delta_{ij} P_j, \quad (\text{no sum on } j), \quad (\text{A10.22})$$

and satisfy the *completeness relation*

$$\sum_{i=1}^r P_i = \mathbf{1}. \quad (\text{A10.23})$$

As $\text{tr}(\mathbf{C} P_i \mathbf{C}^\dagger) = \text{tr } P_i$, the dimension of the i th subspace is given by

$$d_i = \text{tr } P_i. \quad (\text{A10.24})$$

It follows from the characteristic equation (A10.19) and the form of the projection operator (A10.20) that λ_i is the eigenvalue of $\mathbf{M}$ on P_i subspace:

$$\mathbf{M} P_i = \lambda_i P_i, \quad (\text{no sum on } i). \quad (\text{A10.25})$$

Hence, any matrix polynomial $f(\mathbf{M})$ takes the scalar value $f(\lambda_i)$ on the P_i subspace

$$f(\mathbf{M}) P_i = f(\lambda_i) P_i. \quad (\text{A10.26})$$

This, of course, is the reason why one wants to work with irreducible reps: they reduce matrices and “operators” to pure numbers.

A10.2.2 Irreducible representations

Suppose there exist several linearly independent invariant $[d \times d]$ hermitian matrices $\mathbf{M}_1, \mathbf{M}_2, \dots$, and that we have used $\mathbf{M}_1$ to decompose the d -dimensional vector space $V = V_1 \oplus V_2 \oplus \dots$. Can $\mathbf{M}_2, \mathbf{M}_3, \dots$ be used to further decompose V_i ? Further decomposition is possible if, and only if, the invariant matrices commute:

$$[\mathbf{M}_1, \mathbf{M}_2] = 0, \quad (\text{A10.27})$$

or, equivalently, if projection operators P_j constructed from $\mathbf{M}_2$ commute with projection operators P_i constructed from $\mathbf{M}_1$,

$$P_i P_j = P_j P_i. \quad (\text{A10.28})$$

Usually the simplest choices of independent invariant matrices do not commute. In that case, the projection operators P_i constructed from $\mathbf{M}_1$ can be used to project commuting pieces of $\mathbf{M}_2$:

$$\mathbf{M}_2^{(i)} = P_i \mathbf{M}_2 P_i, \quad (\text{no sum on } i).$$

That $\mathbf{M}_2^{(i)}$ commutes with $\mathbf{M}_1$ follows from the orthogonality of P_i :

$$[\mathbf{M}_2^{(i)}, \mathbf{M}_1] = \sum_j \lambda_j [\mathbf{M}_2^{(i)}, P_j] = 0. \quad (\text{A10.29})$$

Now the characteristic equation for $\mathbf{M}_2^{(i)}$ (if nontrivial) can be used to decompose V_i subspace.

An invariant matrix $\mathbf{M}$ induces a decomposition only if its diagonalized form (A10.18) has more than one distinct eigenvalue; otherwise it is proportional to the unit matrix and commutes trivially with all group elements. A rep is said to be *irreducible* if all invariant matrices that can be constructed are proportional to the unit matrix.

According to (A10.13), an invariant matrix $\mathbf{M}$ commutes with group transformations $[G, \mathbf{M}] = 0$. Projection operators (A10.20) constructed from $\mathbf{M}$ are polynomials in $\mathbf{M}$, so they also commute with all $g \in \mathcal{G}$:

$$[G, P_i] = 0 \quad (\text{A10.30})$$

Hence, a $[d \times d]$ matrix rep can be written as a direct sum of $[d_i \times d_i]$ matrix reps:

$$G = \mathbf{1} G \mathbf{1} = \sum_{i,j} P_i G P_j = \sum_i P_i G P_i = \sum_i G_i. \quad (\text{A10.31})$$

In the diagonalized rep (A10.21), the matrix $\mathbf{g}$ has a block diagonal form:

$$\mathbf{C} \mathbf{g} \mathbf{C}^\dagger = \begin{bmatrix} \mathbf{g}_1 & 0 & 0 \\ 0 & \mathbf{g}_2 & 0 \\ 0 & 0 & \ddots \end{bmatrix}, \quad \mathbf{g} = \sum_i \mathbf{C}^i \mathbf{g}_i \mathbf{C}_i. \quad (\text{A10.32})$$

The rep $\mathbf{g}_i$ acts only on the d_i -dimensional subspace V_i consisting of vectors $P_i q$, $q \in V$. In this way an invariant $[d \times d]$ hermitian matrix $\mathbf{M}$ with r distinct eigenvalues induces a decomposition of a d -dimensional vector space V into a direct sum of d_i -dimensional vector subspaces V_i :

$$V \xrightarrow{\mathbf{M}} V_1 \oplus V_2 \oplus \dots \oplus V_r. \quad (\text{A10.33})$$

A10.3 Hénon map symmetries

We note here a few simple symmetries of the Hénon map (3.18). For $b \neq 0$ the Hénon map is reversible: the backward iteration of (3.19) is given by

$$x_{n-1} = -\frac{1}{b}(1 - ax_n^2 - x_{n+1}). \quad (\text{A10.34})$$

Hence the time reversal amounts to $b \rightarrow 1/b$, $a \rightarrow a/b^2$ symmetry in the parameter plane, together with $x \rightarrow -x/b$ in the coordinate plane, and there is no need to explore the (a, b) parameter plane outside the strip $b \in \{-1, 1\}$. For $b = -1$ the map is orientation and area preserving ,

$$x_{n-1} = 1 - ax_n^2 - x_{n+1}, \quad (\text{A10.35})$$

the backward and the forward iteration are the same, and the non-wandering set is symmetric across the $x_{n+1} = x_n$ diagonal. This is one of the simplest models of a return map for a Hamiltonian flow. For the orientation reversing $b = 1$ case we have

$$x_{n-1} = 1 - ax_n^2 + x_{n+1}, \quad (\text{A10.36})$$

and the non-wandering set is symmetric across the $x_{n+1} = -x_n$ diagonal.

References

- [1] W. G. Harter, *Principles of Symmetry, Dynamics, and Spectroscopy* (Wiley, New York, 1993).

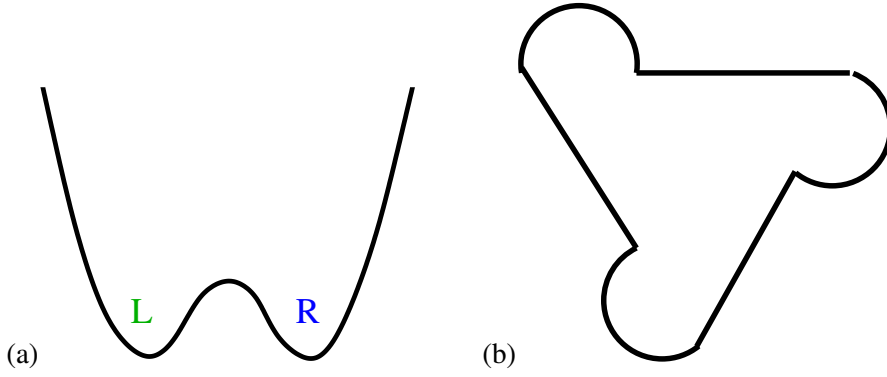


Figure A10.1: (a) A reflection-symmetric double-well potential. (b) A cyclic group C_3 (1/3rd-circle rotation) symmetric plane billiard (infinite wall potential in 2D). (B. Gutkin)

A10.4 Examples

Example A10.1. Reflection and discrete rotation symmetries:

- (a) Reflection symmetric potential $V(x) = sV(x) = V(-x)$:

$$\left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x)\right)\psi(x) = E_n\psi(x) \quad (\text{A10.37})$$

(see figure A10.1). If $\psi(x)$ is solution then $s\psi(x)$ is also solution. From this and non-degeneracy of the spectrum follows that either $s\psi(x) = \psi(x)$ or $s\psi(x) = -\psi(x)$. The first case corresponds to symmetric functions while the second one to anti-symmetric one. Thus the whole spectrum can be decomposed in accordance to a symmetry of the Hamiltonian (equations of motion).

- (b) C_3 rotation symmetry $V(x) = rV(x)$, $G = \{e, r, r_2\}$: By the same argument we have three possibilities:

$$r\psi(x) = \psi(x); \quad r\psi(x) = e^{i2\pi/3}\psi(x); \quad r_{-1}\psi(x) = e^{-i2\pi/3}\psi(x).$$

In addition, by the time reversal symmetry if $\psi(x)$ is solution then $\psi^*(x)$ is solution with the same eigenvalue as well. From this follows that the spectrum must be degenerate. The spectrum is split into a real eigenfunction $\{\psi_1(x)\}$, and a degenerate pair of real eigenfunctions

$$\psi_2(x) = \psi(x) + \psi^*(x); \psi_3(x) = i(\psi(x) - \psi^*(x)), \quad \text{where } r\psi(x) = e^{i2\pi/3}\psi(x)$$

invariant under rotations by 1/3-rd of a circle.

(B. Gutkin)

Example A10.2. Vibrational spectra of molecules: In the linear, harmonic oscillator approximation the deterministic dynamics of the molecule is governed by the Hamiltonian

$$H = \sum_{i=1}^N \frac{m_i}{2} \dot{x}_i^2 + \frac{1}{2} \sum_{i,j=1}^N x_i^\top V_{ij} x_j,$$

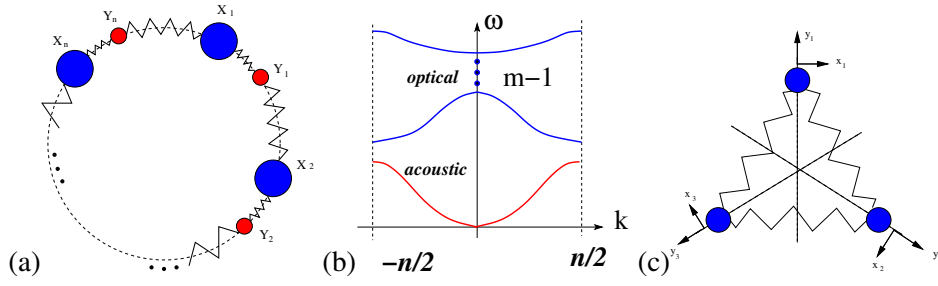


Figure A10.2: (a) Chain with C_3 cyclic symmetry. (b) Dependence of frequency ω on the representation wavenumber k . (c) A molecule with D_3 dihedral symmetry. (B. Gutkin)

where $\{x_i\}$ are small deviations from the resting the equilibrium, resting points of the molecules labelled i . V_{ij} is a symmetric matrix, so it can be brought to a diagonal form by an orthogonal transformation, a set of N uncoupled harmonic oscillators or *normal modes* of frequencies $\{\omega_i\}$.

$$x \rightarrow y = Ux, \quad H = \sum_{i=1}^N \frac{m_i}{2} (\dot{y}_i^2 + \omega_i^2 y_i^2). \quad (\text{A10.38})$$

Consider now the ring of pair-wise interactions of two kinds of molecules sketched in figure A10.2 (a), given by the potential

$$V(z) = \frac{1}{2} \sum_{i=1}^N (k_1(x_i - y_i)^2 + k_2(x_{i+1} - y_i)^2), \quad z_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix}, \quad (\text{A10.39})$$

whose $[2N \times 2N]$ matrix form is (aside to the cognoscenti: this is a Toeplitz matrix):

$$V_{ij} = \frac{1}{2} \begin{pmatrix} k_1 + k_2 & -k_1 & 0 & 0 & 0 & \dots & 0 & 0 & -k_2 \\ -k_1 & k_1 + k_2 & -k_2 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & -k_2 & k_1 + k_2 & -k_1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & -k_1 & k_1 + k_2 & -k_2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & -k_2 & k_1 + k_2 & -k_1 \\ -k_2 & 0 & 0 & 0 & 0 & \dots & 0 & -k_1 & k_1 + k_2 \end{pmatrix}$$

This potential matrix is a holy mess. How do we find an orthogonal transformation (A10.38) that diagonalizes it? Look at figure A10.2 (a). Molecules lie on a circle, so that suggests we should use a Fourier representation. As the $i = 1$ labelling of the starting molecule on a ring is arbitrary, we are free to relabel them, for example use the next molecule pair as the starting one. This relabelling is accomplished by the $[2N \times 2N]$ permutation matrix (or ‘one-step shift’, ‘stepping’ or ‘translation’ matrix) M of form

$$\underbrace{\begin{pmatrix} 0 & 0 & \dots & 0 & I \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \end{pmatrix}}_M \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_n \end{pmatrix} = \begin{pmatrix} z_n \\ z_1 \\ z_2 \\ \vdots \\ z_{n-1} \end{pmatrix}, \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad z_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix} \quad (\text{A10.40})$$

Projection operators corresponding to M are worked out in example A24.2. They are N

distinct $[2N \times 2N]$ matrices,

$$P_k = \begin{pmatrix} I & \bar{\omega}I & \bar{\omega}^2I & \dots & \bar{\omega}^{N-2}I & \bar{\omega}^{N-1}I \\ \omega I & I & \bar{\omega}I & \dots & \bar{\omega}^{N-3}I & \bar{\omega}^{N-2}I \\ \omega^2I & \omega I & I & \dots & \bar{\omega}^{N-4}I & \bar{\omega}^{N-3}I \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \omega^{N-2}I & \omega^{N-3}I & \omega^{N-4}I & \dots & I & \bar{\omega}I \\ \omega^{N-1}I & \omega^{N-2}I & \omega^{N-2}I & \dots & \omega I & I \end{pmatrix}, \quad \omega = e^{i2\pi/N} \quad (\text{A10.41})$$

which decompose the $2N$ -dimensional configuration space of the molecule ring into a direct sum of N 2-dimensional spaces, one for each discrete Fourier mode $k = 0, 1, 2, \dots, N-1$.

The system (A10.39) is clearly invariant under the cyclic permutation relabelling M , $[V, M] = 0$ (though checking this by explicit matrix multiplications might be a bit tedious), so the P_k decompose the interaction potential V as well, and reduce its action to the k th 2-dimensional subspace. Thus the $[2N \times 2N]$ diagonalization (A10.38) is now reduced to a $[2 \times 2]$ diagonalization which one can do by hand. The resulting k th space is spanned by two $2N$ -dimensional vectors, which we guess to be of form:

$$\eta_1 = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 \\ 0 \\ \omega \\ 0 \\ \vdots \\ \omega^{n-1} \\ 0 \end{pmatrix}, \quad \eta_2 = \frac{1}{\sqrt{n}} \begin{pmatrix} 0 \\ 1 \\ 0 \\ \omega \\ \vdots \\ 0 \\ \omega^{n-1} \end{pmatrix}.$$

In order to find eigenfrequencies we have to consider action of V on these two vectors:

$$V\eta_1 = (k_1 + k_2)\eta_1 - (k_1 + k_2\omega)\eta_2, \quad V\eta_2 = (k_1 + k_2)\eta_2 - (k_1 + k_2\bar{\omega})\eta_1.$$

The corresponding eigenfrequencies w (apologies for the notation) are determined by the equation:

$$0 = \det \left[\begin{pmatrix} k_1 + k_2 & -(k_1 + k_2\omega) \\ -(k_1 + k_2\bar{\omega}) & k_1 + k_2 \end{pmatrix} - \frac{w^2}{2} I \right] \implies \frac{1}{2} w_{\pm}^2(k) = k_1 + k_2 \pm |k_1 + k_2\omega^k|, \quad (\text{A10.42})$$

one acoustic ($w(0) = 0$), one optical, see figure A10.2 (b) and the [acoustic and optical phonons](#) wiki. (B. Gutkin)

A10.5 Exercises

A10.1. **Am I a group?** Show that multiplication table

	e	a	b	c	d	f
e	e	a	b	c	d	f
a	a	e	d	b	f	c
b	b	d	e	f	c	a
c	c	b	f	e	a	d
d	d	f	c	a	e	b
f	f	c	a	d	b	e

describes a group. Or does it? (Hint: check whether this table satisfies the group axioms of appendix A10.1.)

From W.G. Harter [1]

A10.2. **Three coupled pendulums with a C_2 symmetry.** Consider 3 pendulums in a row: the 2 outer ones of the same mass m and length l , the one midway of same length but different mass M , with the tip coupled to the tips of the outer ones with springs of stiffness k . Assume displacements are small, $x_i/l \ll 1$.

(a) Show that the acceleration matrix $\ddot{\mathbf{x}} = -\mathbf{a} \mathbf{x}$ is

$$\begin{bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \\ \ddot{x}_3 \end{bmatrix} = - \begin{bmatrix} a+b & -a & 0 \\ -c & 2c+b & -c \\ 0 & -a & a+b \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

where $a = k/ml$, $c = k/Ml$ and $b = g/l$.

(b) Check that $[\mathbf{a}, \mathbf{r}] = 0$, i.e., that the dynamics is invariant under $C_2 = \{e, r\}$, where $\mathbf{r}$ interchanges the outer pendulums,

$$\mathbf{r} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

(c) Construct the corresponding projection operators P_+ and P_- , and show that the 3-pendulum system decomposes into a 1-dimensional subspace, with eigenvalue $(\omega^{(-)})^2 = a + b$, and a 2-dimensional subspace, with acceleration matrix (trust your own algebra, if it strays from what is stated here)

$$\mathbf{a}^{(+)} = \begin{bmatrix} a+b & -\sqrt{2}a \\ -\sqrt{2}c & c+b \end{bmatrix}.$$

The exercise is simple enough that you can do it without using the symmetry, so: construct P_+ , P_- first, use them to reduce $\mathbf{a}$ to irreps, then proceed with computing remaining eigenvalues of $\mathbf{a}$.

(d) Does anything interesting happen if $M = m$?

The point of the above exercise is that almost always the symmetry reduction is only partial: a matrix representation of dimension d gets reduced to a set of subspaces whose dimensions $d^{(\alpha)}$ satisfy $\sum d^{(\alpha)} = d$. Beyond that, love many, trust few, and paddle your own canoe.

From W.G. Harter [1]

A10.3. Lorenz system in polar coordinates: dynamics. (continuation of exercise 11.4)

1. Show that (11.15) has two equilibria:

$$\begin{aligned}(r_0, z_0) &= (0, 0), & \theta_0 &\text{undefined} \\ (r_1, \theta_1, z_1) &= (\sqrt{2b(\rho-1)}, \pi/4, \rho-1).\end{aligned}\tag{A10.43}$$

2. Verify numerically that the eigenvalues and eigenvectors of the two equilibria are (we list here the precise numbers to help you check your programs):

$EQ_1 = (0, 12, 27)$ **equilibrium:** (and its r -rotation EQ_2) has one stable real eigenvalue $\lambda^{(1)} = -13.854578$,

and the unstable complex conjugate pair

$$\lambda^{(2,3)} = \mu^{(2)} \pm i\omega^{(2)} = 0.093956 \pm i10.194505.$$

The unstable eigenplane is defined by eigenvectors

$$\text{Re } \mathbf{e}^{(2)} = (-0.4955, -0.2010, -0.8450)$$

$$\text{Im } \mathbf{e}^{(2)} = (0.5325, -0.8464, 0)$$

with period $T = 2\pi/\omega^{(2)} = 0.6163306$,

radial expansion multiplier

$$\Lambda_r = \exp(2\pi\mu^{(2)}/\omega^{(2)}) = 1.059617,$$

and the contracting multiplier

$$\Lambda_c = \exp(2\pi\mu^{(1)}/\omega^{(2)}) \approx 1.95686 \times 10^{-4}$$

along the stable eigenvector of EQ_1 ,

$$\mathbf{e}^{(3)} = (0.8557, -0.3298, -0.3988).$$

$EQ_0 = (0, 0, 0)$ **equilibrium:** The stable eigenvector $\mathbf{e}^{(1)} = (0, 0, 1)$ of EQ_0 , has contraction rate

$$\lambda^{(2)} = -b = -2.666 \dots$$

The other stable eigenvector is

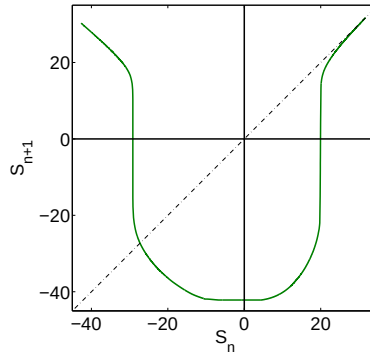
$$\mathbf{e}^{(2)} = (-0.244001, -0.969775, 0), \text{ with contracting eigenvalue}$$

$$\lambda^{(2)} = -22.8277. \text{ The unstable eigenvector}$$

$$\mathbf{e}^{(3)} = (-0.653049, 0.757316, 0) \text{ has eigenvalue}$$

$$\lambda^{(3)} = 11.8277.$$

3. Plot the Lorenz strange attractor both in the Lorenz coordinates figure 2.5, and in the doubled-polar angle coordinates (11.13) for the Lorenz parameter values $\sigma = 10$, $b = 8/3$, $\rho = 28$. Topologically, does it resemble the Lorenz butterfly, the Rössler attractor, or neither? The Poincaré section of the Lorenz flow fixed by the z -axis and the equilibrium in the doubled polar angle representation, and the corresponding return map (s_n, s_{n+1}) are plotted in figure 14.14.
4. Construct the return map (s_n, s_{n+1}) ,



where s is arc-length measured along the unstable manifold of EQ_0 , lower Poincaré section of figure 14.14 (b). Elucidate its relation to the return map of figure 14.15. (plot by J. Halcrow)

5. Show that if a periodic orbit of the polar representation Lorenz is also periodic orbit of the Lorenz flow, their Floquet multipliers are the same. How do the Floquet multipliers of relative periodic orbits of the representations relate to each other?
6. What does the volume contraction formula (4.42) look like now? Interpret.

Appendix A14

Charting the state space

PERIODIC ORBITS of unimodal mappings are studied in sect. A14.1. Pruning theory for Bernoulli shifts is discussed in sect. A14.3.

A14.1 Periodic orbits of unimodal maps

A *periodic point* (*cycle point*) x_k belonging to a cycle of period n is a real solution of

$$f^n(x_k) = f(f(\dots f(x_k) \dots)) = x_k, \quad k = 0, 1, 2, \dots, n-1. \quad (\text{A14.1})$$

The n th iterate of a unimodal map has at most 2^n monotone segments, and therefore there will be 2^n or fewer periodic points of period n . Similarly, the backward and the forward Smale horseshoes intersect at most 2^n times, and therefore there will be 2^n or fewer periodic points of period n . A periodic orbit of period n corresponds to an infinite repetition of a period $n = n_p$ symbol string, customarily indicated by a line over the string:

$$S_p = (s_1 s_2 s_3 \dots s_n)^\infty = \overline{s_1 s_2 s_3 \dots s_n}.$$

As all itineraries are infinite, we shall adopt convention that a finite string itinerary $S_p = s_1 s_2 s_3 \dots s_n$ stands for infinite repetition of a finite block, and routinely omit the overline. x_0 , its cyclic permutation $\overline{s_k s_{k+1} \dots s_n s_1 \dots s_{k-1}}$ corresponds to the point x_{k-1} in the same cycle. A cycle p is called *prime* if its itinerary S cannot be written as a repetition of a shorter block S' .

Here we give explicit formulas for the topological coordinate of a periodic point, given its itinerary. For the purpose of what follows it is convenient to compactify the itineraries by replacing the binary alphabet $s_i = \{0, 1\}$ by the infinite alphabet

$$\{a_1, a_2, a_3, a_4, \dots; \bar{0}\} = \{1, 10, 100, 1000, \dots; \bar{0}\}. \quad (\text{A14.2})$$

In this notation the itinerary $S = a_i a_j a_k a_l \dots$ and the corresponding topological coordinate (14.4) are related by $\gamma(S) = .1^i 0^j 1^k 0^l \dots$. For example:

$$\begin{aligned} S &= 111011101001000\dots = a_1 a_1 a_2 a_1 a_1 a_2 a_3 a_4 \dots \\ \gamma(S) &= .101101001110000\dots = .1^1 0^1 1^2 0^1 1^1 0^2 1^3 0^4 \dots \end{aligned}$$

Cycle points whose itineraries start with $w_1 = w_2 = \dots = w_i = 0$, $w_{i+1} = 1$ remain on the left branch of the tent map for i iterations, and satisfy $\gamma(0\dots 0S) = \gamma(S)/2^i$.

Periodic points correspond to rational values of γ , but we have to distinguish *even* and *odd* cycles. The even (odd) cycles contain even (odd) number of a_i in the repeating block, with periodic points given by

$$\gamma(a_i a_j \dots a_k a_\ell) = \begin{cases} \frac{2^n}{2^n - 1} \cdot 1^i 0^j \dots 1^k & \text{even} \\ \frac{1}{2^n + 1} (1 + 2^n \times 1^i 0^j \dots 1^\ell) & \text{odd,} \end{cases} \quad (\text{A14.3})$$

where $n = i + j + \dots + k + \ell$ is the cycle period. The maximal value periodic point is given by the cyclic permutation of S with the largest a_i as the first symbol, followed by the smallest available a_j as the next symbol, and so on. For example:

$$\begin{aligned} \hat{\gamma}(1) &= \gamma(a_1) = .10101\dots = \overline{.10} = 2/3 \\ \hat{\gamma}(10) &= \gamma(a_2) = .1^2 0^2 \dots = \overline{.1100} = 4/5 \\ \hat{\gamma}(100) &= \gamma(a_3) = .1^3 0^3 \dots = \overline{.111000} = 8/9 \\ \hat{\gamma}(101) &= \gamma(a_2 a_1) = .1^2 0^1 \dots = \overline{.110} = 6/7 \end{aligned}$$

An example of a cycle where only the third symbol determines the maximal value periodic point is

$$\hat{\gamma}(1101110) = \gamma(a_2 a_1 a_2 a_1 a_1) = \overline{.11011010010010} = 100/129.$$

For the full tent map the periodic points are computed in example 14.10. The maximal values of unimodal map cycles up to period 5 are listed in table 14.1.

A14.2 Unimodal map bifurcation sequences

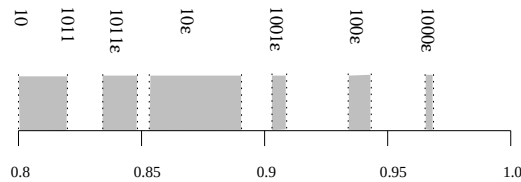
(K.T. Hansen and P. Cvitanović)

Periodic orbits in smooth unimodal maps are generically created either as a pair with one stable and one unstable period n orbit in a saddle node bifurcation point, or as a period $2n$ orbit in a bifurcation where a period n orbit becomes unstable. Immediately after a saddle node bifurcation the two created orbits both have the same itinerary $\overline{s_1 s_2 \dots s_n}$ with an even number of symbols 1 and with the topological parameter value $\kappa(\overline{s_1 s_2 \dots s_n}) = \hat{\gamma}(\overline{s_1 s_2 \dots s_n})$. Orbits with this itinerary exist for all unimodal maps with $\kappa \geq \hat{\gamma}(\overline{s_1 s_2 \dots s_n})$. As the parameter in the smooth unimodal map increases the stable orbit passes a superstable point and changes its symbolic dynamics. If we now assume that the symbol string $\overline{s_1 s_2 \dots s_n}$ is the cyclic permutation giving the maximum γ value, then the itinerary of the stable orbit after the superstable point is $\overline{s_1 s_2 \dots s_{n-1} (1 - s_n)}$, since the point closest to the critical point passes through the critical point. The topological parameter value of the map is then $\kappa(\overline{s_1 s_2 \dots s_{n-1} (1 - s_n)})$. The inadmissible topological parameter interval $(\kappa(\overline{s_1 s_2 \dots s_n}), \kappa(\overline{s_1 s_2 \dots s_{n-1} (1 - s_n)}))$ is then uniquely

S	$\hat{\gamma}(S)$		S	$\hat{\gamma}(S)$	
$\overline{0}$	$\overline{.0}$	$= 0$	$\overline{10111}$	$\overline{.11010}$	$= 26/31$
$\overline{1}$	$\overline{.10}$	$= 2/3$	$\overline{10110}$	$\overline{.1101100100}$	$= 28/33$
$\overline{10}$	$\overline{.1100}$	$= 4/5$	$\overline{10010}$	$\overline{.11100}$	$= 28/31$
$\overline{101}$	$\overline{.1110}$	$= 6/7$	$\overline{10011}$	$\overline{.1110100010}$	$= 10/11$
$\overline{100}$	$\overline{.111000}$	$= 8/9$	$\overline{10001}$	$\overline{.11110}$	$= 30/31$
$\overline{1011}$	$\overline{.11010010}$	$= 14/17$	$\overline{10000}$	$\overline{.1111100000}$	$= 32/33$
$\overline{1001}$	$\overline{.11110}$	$= 14/15$			
$\overline{1000}$	$\overline{.11110000}$	$= 16/17$			

Table A14.1: The maximal values of unimodal map cycles up to period 5. (K.T. Hansen)

Figure A14.1: Bifurcation points from table A14.1 plotted as a function of the topological parameter κ . Gray areas are inadmissible intervals of κ corresponding to stable windows in a smooth unimodal map. As a shorthand notation for pairs of orbits we use the letter ϵ to denote either a 0 or a 1. (K.T. Hansen)



related to the parameter interval in a between the saddle node bifurcation and the superstable point, or more loosely speaking; to the a interval where the orbit $\overline{s_1 s_2 \dots s_{n-1}(1 - s_n)}$ is stable.

In the same way there will be an interval

$$(\overline{\kappa(s_1 s_2 \dots s_{n-1}(1 - s_n))}, \overline{\kappa(s_1 s_2 \dots s_{n-1}(1 - s_n) s_1 s_2 \dots s_n)})$$

corresponding to the interval in a from where the orbit $\overline{s_1 s_2 \dots s_{n-1}(1 - s_n)}$ is superstable to the point where the orbit $\overline{s_1 s_2 \dots s_{n-1}(1 - s_n) s_1 s_2 \dots s_n}$ is superstable. This interval includes the period doubling bifurcation where the $2n$ orbit $\overline{s_1 s_2 \dots s_{n-1}(1 - s_n) s_1 s_2 \dots s_n}$ is created.

From table A14.1 we can find some of the largest intervals in κ corresponding to the stability windows in a smooth unimodal map. The stable period 3 orbit window on the parameter a -axis corresponds to the interval $(6/7, 8/9)$ on the κ line and so on, see figure A14.1.

A14.3 Pruned Bernoulli shift

In this section we illustrate extraction of a symbolic dynamics on a piecewise linear repeller for which the itinerary of a repeller point x is given by its binary expansion. The main result is the Algorithm 1 which converts recursively a given value of the “pruning point” x_p into the symbolic dynamics of the map. We shall apply this symbolic dynamics in the next section to construction of explicit examples of zeta functions.

The simplest example of a map with complete binary dynamics is the Bernoulli shift

$$x_{n+1} = 2x_n \mod 1. \quad (\text{A14.4})$$

The map acts by shifting the binary point to the right, so the itinerary of $x = x_0$ (if $x_n < 1/2$, then $s_n = 0$; if $x_n > 1/2$, then $s_n = 1$) is simply its binary expansion $x = .s_1 s_2 s_3 \dots$. The periodic points $\overline{s_1 s_2 \dots s_n}$ correspond to rational x :

$$x_{\overline{s_1 s_2 \dots s_n}} = \sum_{k=1}^n \frac{s_k}{2^k} \sum_{m=0}^{\infty} \frac{1}{2^{nm}} = \frac{2^n}{2^n - 1} .s_1 s_2 \dots s_n = \frac{\sum_{k=1}^n s_k 2^{n-k}}{2^n - 1}. \quad (\text{A14.5})$$

This is just the binary version of the familiar fact that the decimal expansion of a rational number is (eventually) periodic.

The *clipped Bernoulli shift* obtained by slicing off all $x \geq x_p$ is a two-branch linear map of form

$$\begin{aligned} f_0(x) &= 2x & 0 \leq x \leq 1/2 \\ f_1(x) &= 2x - 1 & 1/2 \leq x < x_p. \end{aligned} \quad (\text{A14.6})$$

We shall refer to x_p as the *pruning point*. All trajectories that land in the $x_p < x \leq 1$ interval escape; only those x whose binary expansion contains no subsequence $_{s_m s_{m+1} s_{m+2} \dots s_{m+n}}$ such that $.s_m s_{m+1} s_{m+2} \dots s_{m+n} > x_p$ survive the pruning. The surviving trajectories are unstable (with the Floquet multiplier $\Lambda_c = 2^{n_c}$, where n_c is the length of the trajectory c), and form a repelling strange set. The symbolic dynamics of the clipped Bernoulli shift is specified by the binary expansion of x_p , in the same sense that the symbolic dynamics of a unimodal map is specified by the trajectory (the kneading invariant [3]) of its critical value $x_p = f(x_c)$.

We find this map pedagogically convenient, as the technique for extracting the symbolic dynamics is essentially the same as for the unimodal maps of sect. A14.2, but the conversion of the parameter x_p into symbolic dynamics is somewhat simpler, as the itineraries are ordered monotonically by the ordinary binary tree rather than the alternating binary tree.

Our strategy for converting x_p into symbolic dynamics is to check recursively whether the x_p falls into a window at successive levels of resolution. If it does, we obtain the exact alphabet; if it does not, the last letter in the approximate alphabet has to be refined, and the procedure repeated. We first phrase this recursive procedure as a general algorithm and then illustrate it by a few examples (the reader might prefer to glance at those first).

Algorithm 1: Pruning the symbolic dynamics from above

Given the pruning point x_p , the symbolic dynamics is determined recursively as follows:

1. expand $x_p = .s_1 s_2 s_3 \dots$ in binary $s_i = \{0, 1\}$.

2. compactify the binary symbol sequences by rewriting them in the alphabet $\{a_1, a_2, a_3, a_4, \dots\} = \{0, \underline{10}, \underline{110}, \underline{1110}, \dots\}$ where a_n stands for a block of $n-1$ binary 1's followed by 0. At this level of resolution x_p is bracketed by

$$.a_n < x_p \leq .a_{n+1}$$

and the approximate alphabet is $\{a_1, a_2, a_3, a_4, \dots, a_n\}$.

2. prohibition of a_{n+1} implies that the rightmost surviving point is $.\overline{a_n}$. If the pruning point x_p is in the window $\Delta_{a_n} = (.\overline{a_n}, .a_{n+1}]$, the exact alphabet is $\{a_1, a_2, a_3, \dots, a_n\}$, and the algorithm stops.

3. If $x_p < .\overline{a_n}$, the pruning point falls somewhere within the $l_{a_n} = (.a_n, .\overline{a_n}]$ interval, and not all sequences starting with $.a_n$ are allowed. Subdivide the $.a_n$ interval into finer subintervals by appending to $.a_n$ all allowed basic blocks: $\{a_n\} = \{a_n a_1, a_n a_2, a_n a_3, \dots, a_n a_m\}$. Here $m \leq n$ is determined by the value of x_p , $.a_n a_m < x_p \leq .a_n a_{m+1}$. The refined approximate alphabet is given by $\{a_1, a_2, a_3, \dots, a_{n-1}, b_1, b_2, \dots, b_m\} = \{a_1, a_2, a_3, \dots, a_{n-1}, \underline{a_n a_1}, \underline{a_n a_2}, \dots, \underline{a_n a_m}\}$.

4. repeat step 2: if $x_p \in \Delta_{b_m} = (.b_m, .\overline{b_{m+1}}]$, the above alphabet is exact, otherwise continue refining the $l_{b_m} = (.b_m, .\overline{b_m}]$ interval by replacing b_m by new letters $\{c_1, c_2, \dots, c_k\}$, as in step 3.

Clearly any x_p whose binary expansion is finite yields a finite alphabet, and so does any x_p that falls into a Δ_{c_n} window. Otherwise the algorithm generates a monotone sequence of $l_{c_n} = (.c_n, .\overline{c_n}]$ covering intervals, $x_p \in l_{c_n}$, together with the associated approximate alphabets.

Example 1: the ‘golden mean’ pruning $x_p = .11$

The simplest example of pruning for the clipped Bernoulli shift is given by the $x_p = .11$ pruning point value. As the substring $\_11\_$ is forbidden, 1 must always be followed by 0, so the allowed sequences can be built from any number of consecutive 0's and 10 blocks, and the alphabet is simply $\{0, 10\}$. Note that if x_p is set to $.11$, the rightmost surviving point of the repeller is not $.11$, but $.10101010\dots$, *ie.* the periodic point $.\overline{10}$. Hence any x_p value in the window $\Delta_{10} = (.10, .11] = (2/3, 3/4]$ leads to the same symbolic dynamics.

Example 2: The right ascending staircase $\Delta_0, \Delta_{10}, \Delta_{110}, \dots$,

By the same argument as the above, $x_p = .11\dots 1$ (n binary “1”s, followed by a “0”) pruning leads to the n letter alphabet

$$\{a_1, a_2, a_3, \dots, a_n\} = \{0, 10, 110, 1110, \dots, 11\dots 10\},$$

and the symbolic dynamics unchanged over windows

$$\Delta_{11\dots 10} = (.11\dots 10, .11\dots 1] = \left(\frac{2^{n-1} - 2}{2^n - 1}, \frac{2^n - 1}{2^n} \right], \quad (\text{A14.7})$$

whose width is shrinking as $|\Delta_{11\dots 10}| = \frac{1}{2^n(2^n-1)}$.

Example 3: a “typical” pruning front value: $x_p = .11101101110$

1. At the first level of resolution x_p is bracketed by

$$.111 < x_p \leq .1111$$

so 1 can appear only within blocks 10, 110 and 1110. Rewrite x_p in the new alphabet $\{a_1, a_2, a_3, a_4\} = \{0, \underline{10}, \underline{110}, \underline{1110}\}$, where a_n stands for a block of $n-1$ binary 1’s followed by 0: $x_p = .a_4a_3a_4a_2$

2. As in the above example, prohibition of $a_5 = 11110$ implies that the rightmost surviving point is $\overline{.a_4} = .\underline{1110}$. Were $x_p \in \Delta_{a_4} = (\overline{.a_4}, .a_5]$, the alphabet $\{a_1, a_2, a_3, a_4\}$ would be exact, and we would be finished. However, as $x_p < \overline{.a_4}$, the pruning point falls somewhere within $.a_4 < x_p \leq \overline{.a_4}$, and not all sequences starting with a_4 are allowed.

3. Therefore we subdivide the $.1110$ interval into finer subintervals by appending to $.a_4$ all allowed basic blocks: $\{a_4\} = \{a_4a_1, a_4a_2, a_4a_3, a_4a_4\}$. As $x_p = .a_4a_3a_4a_2 < .a_4a_4$, the $.a_4a_4$ interval is pruned. x_p can be rewritten in the new alphabet $\{a_1, a_2, a_3, b_1, b_2, b_3\} = \{a_1, a_2, a_3, \underline{a_4a_1}, \underline{a_4a_2}, \underline{a_4a_3}\}$ as $x_p = .b_3b_2$.

4. repeat step 2: is $x_p > \overline{.b_3}$? It is not, so

5. repeat step 3: subdivide $\{b_3\} = \{b_3a_1, b_3a_2, b_3a_3, b_3b_1, b_3b_2\}$. The b_3b_2 block is forbidden by the pruning point value, so we are done; the alphabet consists of 9 letters

$$\{a_1, a_2, a_3, b_1, b_2, c_1, c_2, c_3, c_4\} = \{a_1, a_2, a_3, b_1, b_2, \underline{b_3a_1}, \underline{b_3a_2}, \underline{b_3a_3}, \underline{b_3b_1}\}$$

We could have kept the binary notation throughout, but a two-letter alphabet makes for rather tedious reading; in the binary notation the fundamental blocks are

$$\{0, 10, 110, 11100, 111010, 11101100, 111011010, 1110110110, 111011011100\}$$

.

This finishes the list of examples.

For the clipped Bernoulli map the fraction of the x_p parameter values for which the alphabet is finite can be estimated analytically. If c is a sequence corresponding to one of the windows unfolded recursively in the above, the symbolic dynamics is unchanged over the window

$$\Delta_c = (\overline{.c}, .c1] = \left(\frac{.c}{1 - 2^{-n_c}}, .c1\right] \quad (\text{A14.8})$$

whose width shrinks with n_c , the length of the binary string c , as

$$|\Delta_c| = .c1 - \overline{.c} = \frac{1 - .c_{n+1}}{2^{n_c} - 1} = \frac{.c^+ 1}{2^{n_c} - 1}, \quad (\text{A14.9})$$

where c^+ is the binary complement of c . This follows from summing the successive images of the pruned interval $1 - .c1$ within the $(.c, .c1)$ interval.

The widths of the fatest $c = .1000 \cdots 0$ and the thinnest $c = .11 \cdots 1$ steps in the devil staircase corresponding to strings of length n are, respectively,

$$\frac{2^{n-1} - 1}{2^n(2^n - 1)} \geq \Delta_c \geq \frac{1}{2^n(2^n - 1)} .$$

The lower bound follows from (A14.7) and the upper bound from $1 - .c1 = 1 - .10 \cdots 01 = 1/2 - 1/2^n$.

We leave the evaluation of the total measure $\sum \Delta_c$ taken up by finite alphabets as an exercise for the student. What is the measure taken up by the infinite alphabets? Is there a set of non-integer Hausdorff dimension, and what is its significance?

A14.3.1 Topological entropy

The symbolic dynamics considered in the preceeding section gives a class of rather simple topological polynomials. If the symbolic dynamics can be written as a complete (unrestricted, unpruned) alphabet in N symbols, then $t_f = 1$ if $f \in$ alphabet, $t_f = 0$ otherwise. According to the results of the preceeding section, for the clipped Bernoulli shift the symbolic dynamics is given by a finite (or infinite alphabet) built up of blocks of increasing binary length:

$$\{a_1, a_2, a_3, \cdots, a_n, b_1, b_2, \cdots, b_m, c_1, c_2, \cdots\}$$

For a finite unrestricted alphabet, the topological entropy is given by the smallest root of the corresponding topological polynomial:

$$\begin{aligned} 0 = & 1 - z^{na_1} - z^{na_2} - z^{na_3} - \cdots - z^{na_n} \\ & - z^{nb_1} - z^{nb_2} - \cdots - z^{nb_m} \\ & - z^{nc_1} - z^{nc_2} - \cdots - z^{nc_k} . \end{aligned} \quad (\text{A14.10})$$

The simplest example of a not entirely trivial topological polynomial follows from the 3-cycle pruning example 1. The fundamental cycles $0, \underline{01}$ are of period 1 and 2, so the topological polynomial is simply

$$\prod_p (1 - z^{np}) = 1 - z - z^2 , \quad (\text{A14.11})$$

and the topological entropy is $h = \log \frac{1+\sqrt{5}}{2}$.

The topological polynomial for the example 2. is given by

$$\prod_p (1 - z^{np}) = 1 - z - z^2 - \cdots - z^n = \frac{1 - 2z + z^{n+1}}{1 - z} . \quad (\text{A14.12})$$

The topological entropy is $h = \log \lambda_0$, where λ_0 is the leading eigenvalue $1 < \lambda_0 \leq 2$. The remaining roots of (A14.12) lie (for large n) close to the unit circle in the complex plane and are of no physical interest.

The alphabet above was generated by resolving the longest fundamental string at a given level by a set of longer strings; so even if the grammar is not finite and the cycle expansion is not a polynomial, the convergence of the cycle expansion should be good, as the errors are bounded from below and above by truncating the expansions with terms z^n and z^{n+1} , where n is the length of the longest binary string in the alphabet. With increasing resolution n typically grows in leaps and bounds. The entropy is given by the isolated real zero $1/2 \leq z < 1$; the remaining zeros of the polynomial approximations to the entropy function for infinite grammar bunch on the unit circle. The rate of convergence depends on the separation of the leading, entropy eigenvalue from the non-leading eigenvalues; as long as there is a gap, the convergence will be exponential, though situations without gap also arise and are interesting (cf. period doubling $1/\zeta$ in ref. [3]).

The above considerations, in spite of the restriction to mere cycle counting, reveal a great deal about the spectra of more general transfer operators. For linear systems with a single scale Λ , $1/\zeta_0(z)$ is given by (A14.10), simply by rescaling $z \rightarrow z/|\Lambda|$. For nonlinear mappings, polynomial approximations to $1/\zeta_k$ have a rather similar structure; there is a physically significant $\lambda_0^{(k)}$, together with a family of unphysical poles in the complex plane, placed roughly on a circle of radius $|1/c|$, where c controls the asymptotic behavior of curvatures, $c_n \approx c^n$. The extraneous zeros delineate the boundary of the convergence of the cycle expansion of $1/\zeta_k$; and for longer and longer truncated Selberg products $1/\zeta_0 \zeta_1 \cdots \zeta_k$ this boundary is pushed further and further out, allowing determination of a finite number of leading eigenvalues of $\mathcal{L}$.

Commentary

Remark A14.1. What are Manning's multiples? According to Viviane Baladi, the *Red Book* [4], Proposition 2.4 explains Manning's argument to count periodic points. The idea is that you have to be careful with the boundary of the Markov partition and all of its iterates, as the preimages of the boundary are everywhere dense. Manning's paper [2] is explained in a very un-Bourbaki way in Bowen [1], middle of page 14. This is pure combinatorics and $x_{i_0, \dots, i_{m-1}}^*$ is simply a point in $\mathcal{M}$. If you agree that $x_{i_0, \dots, i_{m-1}}^*$ is simply a way to name a point x in $\mathcal{M}$, then you should not be surprised that $T_x^* f^{-m}$ of *Red Book*, Proposition 7.1 denotes the linear bundle map over f^m on the cotangent bundle $T^*\mathcal{M}$. (This map already appeared in Proposition 6.2 with its ℓ -forms brothers and sisters).

In words that mere mortals have a chance of understanding, if $\mathcal{M}_i \cap \mathcal{M}_j \neq \emptyset$, you are double-counting the border points. Therefore you first count all periodic points in $\{\mathcal{M}_0, \mathcal{M}_1, \dots, \mathcal{M}_{m-1}\}$, then subtract all double counts in all pairs of border overlaps $\mathcal{M}_i \cap \mathcal{M}_j$, then add all triple counts in $\mathcal{M}_i \cap \mathcal{M}_j \cap \mathcal{M}_k$ 3-tuples, and so on.

References

- [1] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms* (Springer, New York, 1975).
- [2] A. Manning, “Axiom A diffeomorphisms have rational zeta function”, *Bull. London Math. Soc.* **3**, 215–220 (1971).
- [3] J. Milnor and W. Thurston, “On iterated maps of the interval”, in *Dynamical Systems*, edited by J. C. Alexander (Springer, New York, 1988), pp. 465–563.
- [4] D. Ruelle, *Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics*, 2nd ed. (Cambridge Univ. Press, Cambridge UK, 2004).

Appendix A16

Finding cycles

(C. Chandre)

A16.1 Newton-Raphson method

A16.1.1 Contraction rate

CONSIDER A d -DIMENSIONAL MAP $x' = f(x)$ with an unstable fixed point x_* . The Newton-Raphson algorithm is obtained by iterating the following map

$$x' = g(x) = x - (J(x) - \mathbf{1})^{-1} (f(x) - x).$$

The linearization of g near x_* leads to

$$x_* + \epsilon' = x_* + \epsilon - (J(x_*) - \mathbf{1})^{-1} (f(x_*) + J(x_*)\epsilon - x_* - \epsilon) + O(\|\epsilon\|^2),$$

where $\epsilon = x - x_*$. Therefore,

$$x' - x_* = O((x - x_*)^2).$$

After n steps and if the initial guess x_0 is close to x_* , the error decreases super-exponentially

$$g^n(x_0) - x_* = O((x_0 - x_*)^{2^n}).$$

A16.1.2 Computation of the inverse

The Newton-Raphson method for finding n -cycles of d -dimensional mappings using the multi-shooting method reduces to the following equation

$$\begin{pmatrix} \mathbf{1} & & & -Df(x_n) \\ -Df(x_1) & \mathbf{1} & & \\ & \dots & \mathbf{1} & \\ & & -Df(x_{n-1}) & \mathbf{1} \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \\ \dots \\ \delta_n \end{pmatrix} = - \begin{pmatrix} F_1 \\ F_2 \\ \dots \\ F_n \end{pmatrix}, \quad (\text{A16.1})$$

where $Df(x)$ is the $[d \times d]$ Jacobian matrix of the map evaluated at the point x , and $\delta_m = x'_m - x_m$ and $F_m = x_m - f(x_{m-1})$ are d -dimensional vectors. By some straightforward algebra, the vectors δ_m are expressed as functions of the vectors F_m :

$$\delta_m = - \sum_{k=1}^m \beta_{k,m-1} F_k - \beta_{1,m-1} (\mathbf{1} - \beta_{1,n})^{-1} \left(\sum_{k=1}^n \beta_{k,n} F_k \right), \quad (\text{A16.2})$$

for $m = 1, \dots, n$, where $\beta_{k,m} = Df(x_m)Df(x_{m-1}) \cdots Df(x_k)$ for $k < m$ and $\beta_{k,m} = \mathbf{1}$ for $k \geq m$. Therefore, finding n -cycles by a Newton-Raphson method with multiple shooting requires inversion of a $[d \times d]$ matrix $\mathbf{1} - Df(x_n)Df(x_{n-1}) \cdots Df(x_1)$.

A16.2 Hybrid Newton-Raphson / relaxation method



Consider a d -dimensional map $x' = f(x)$ with an unstable fixed point x_* . The transformed map is the following one:

$$x' = g(x) = x + \gamma C(f(x) - x),$$

where $\gamma > 0$ and C is a $[d \times d]$ invertible constant matrix. We note that x_* is also a fixed point of g . Consider the stability matrix at the fixed point x_*

$$A_g = \left. \frac{dg}{dx} \right|_{x=x_*} = \mathbf{1} + \gamma C(A_f - \mathbf{1}).$$

The matrix C is constructed such that the eigenvalues of A_g are of modulus less than one. Assume that A_f is diagonalizable: In the basis of diagonalization, the matrix writes:

$$\tilde{A}_g = \mathbf{1} + \gamma \tilde{C}(\tilde{A}_f - \mathbf{1}),$$

where $\tilde{A}_f$ is diagonal with elements μ_i . We restrict the set of matrices $\tilde{C}$ to diagonal matrices with $\tilde{C}_{ii} = \epsilon_i$ where $\epsilon_i = \pm 1$. Thus $\tilde{A}_g$ is diagonal with eigenvalues $\gamma_i = 1 + \gamma \epsilon_i (\mu_i - 1)$. The choice of γ and ϵ_i is such that $|\gamma_i| < 1$. It is easy to see that if $\text{Re}(\mu_i) < 1$ one has to choose $\epsilon_i = 1$, and if $\text{Re}(\mu_i) > 1$, $\epsilon_i = -1$. If λ is chosen such that

$$0 < \gamma < \min_{i=1, \dots, d} \frac{2|\text{Re}(\mu_i) - 1|}{|\mu_i - 1|^2},$$

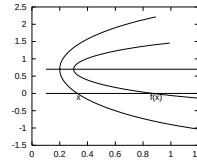
all the eigenvalues of A_g have modulus less than one. The contraction rate at the fixed point for the map g is then $\max_i |1 + \gamma \epsilon_i (\mu_i - 1)|$. If $\text{Re}(\mu_i) = 1$, it is not possible to stabilize x_* by the set of matrices γC .

From the construction of C , we see that 2^d choices of matrices are possible. For example, for 2-dimensional systems, these matrices are

$$C \in \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \right\}.$$

For 2-dimensional dissipative maps, the eigenvalues satisfy $\text{Re}(\mu_1)\text{Re}(\mu_2) \leq \det Df < 1$. The case $(\text{Re}(\mu_1) > 1, \text{Re}(\mu_2) > 1)$ which is stabilized by $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ has to be discarded. The minimal set is reduced to three matrices.

Figure A16.1: Illustration of the optimal Poincaré section. The original section $y = 0$ yields a large distance $x - f(x)$ for the Newton iteration. A much better choice is $y = 0.7$.



A16.2.1 Newton method with optimal section

(F. Christiansen)



In some systems it might be hard to find a good starting guess for a fixed point. This can happen, for example, if the topology and/or the symbolic dynamics of the flow is not well understood. By changing the Poincaré section one might get a better initial guess in the sense that x and $f(x)$ are closer together. We illustrate this in figure A16.1. The figure shows a Poincaré section, $y = 0$, an initial guess x , the corresponding $f(x)$ and pieces of the trajectory near these two points.

If Newton iteration does not converge for the initial guess x we might have to work very hard to find a better guess, particularly if this is in a high-dimensional system (high-dimensional in this context might mean a Hamiltonian system with 3 or more degrees of freedom). Clearly, we could easily obtain a much better guess by simply shifting the Poincaré section to $y = 0.7$ where the distance $x - f(x)$ would be much smaller. Naturally, one cannot so easily determine by inspection the best section for a higher dimensional system. Rather, the way to proceed is as follows: We want to have a minimal distance between our initial guess x and its image $f(x)$. We therefore integrate the flow looking for a minimum in the distance $d(t) = |f^t(x) - x|$. $d(t)$ is now a minimum with respect to variations in $f^t(x)$, but not necessarily with respect to x . We therefore integrate x either forward or backward in time. Doing this minimizes d with respect to x , but now it is no longer minimal with respect to $f^t(x)$. We therefore repeat the steps, alternating between correcting x and $f^t(x)$. In most cases this process converges quite rapidly. The result is a trajectory for which the vector $(f(x) - x)$ connecting the two end points is perpendicular to the flow at both points. We can now define a Poincaré section as the hyper-plane that goes through x and is normal to the flow at x , $(x' - x) \cdot v(x) = 0$.

The image $f(x)$ lies in the section. This section is optimal in the sense that a close return on the section is a local minimum of the distance between x and $f^t(x)$. More important, the part of the stability matrix that describes linearization perpendicular to the flow is exactly the stability of the flow in the section when

$f(x)$ is close to x . With this method, the Poincaré section changes with each Newton iteration. Should we later want to put the fixed point on a specific Poincaré section, it will only be a matter of moving along the trajectory.

Appendix A18

Counting

I've come to the conclusion that god could never understand calculus, or indeed the real numbers. But I am pretty sure that god can count.

— David Finkelstein to Lee Smolin (1972)

THE GENERATING FUNCTIONS of sects. 18.2 and A20.1 are really Laplace transforms in disguise, or Z-transforms. Kneading theory for unimodal mappings is developed in sect. A18.1. The prime factorization for dynamical itineraries of sect. A18.2 illustrates the sense in which prime cycles are “prime” - the product structure of zeta functions is a consequence of the unique factorization property of symbol sequences.

A18.1 Topological zeta functions for infinite subshifts

(P. Dahlqvist)



The transition graph methods outlined in chapter 14 are well suited for symbolic dynamics of finite subshift type. A sequence of well defined rules leads to the answer, the topological zeta function, which turns out to be a polynomial. For infinite subshifts one would have to go through an infinite sequence of graph constructions and it is of course very difficult to make any asymptotic statements about the outcome. Luckily, for some simple systems the goal can be reached by much simpler means. This is the case for unimodal maps.

We will restrict our attention to the topological zeta function for unimodal maps with one external parameter $f_\Lambda(x) = \Lambda g(x)$. As usual, symbolic dynamics is introduced by mapping a time series $\dots x_{i-1} x_i x_{i+1} \dots$ onto a sequence of symbols

$I(C)$	$\zeta_{top}^{-1}(z)/(1-z)$	$I(C)$	$\zeta_{top}^{-1}(z)/(1-z)$
1C	$\prod_{n=0}^{\infty} (1 - z^{2^n})$	1001C	$(1 - 2z^2)/(1 + z)$
101C		100111C	
1011101C		10011C	
$H^{\infty}(1)$		100110C	
10111C		100C	
1011111C		100010C	
101^{∞}		10001C	
101111111C		100011C	
101111C		1000C	
1011C		100001C	
101101C	$(1 - z - z^2)$	10000C	$(1 - 2z)/(1 - z)$
10C		100000C	
10010C		10^{∞}	
100101C			

Table A18.1: All ordered kneading sequences up to length seven, as well as some longer kneading sequences. Harmonic extension $H^{\infty}(1)$ is defined below.

$\dots s_{i-1} s_i s_{i+1} \dots$ where

$$\begin{aligned}
 s_i &= 0 & x_i < x_c \\
 s_i &= C & x_i = x_c \\
 s_i &= 1 & x_i > x_c
 \end{aligned} \tag{A18.1}$$

and x_c is the critical point of the map (i.e., maximum of g). In addition to the usual binary alphabet we have added a symbol C for the critical point. The kneading sequence K_{Λ} is the itinerary of the critical point (14.5). The crucial observation is that no periodic orbit can have a topological coordinate (see sect. A14.1) beyond that of the kneading sequence. The kneading sequence thus inserts a border in the list of periodic orbits (ordered according to maximal topological coordinate), cycles up to this limit are allowed, all beyond are pruned. All unimodal maps (obeying some further constraints) with the same kneading sequence thus have the same set of periodic orbits and the same topological zeta function. The topological coordinate of the kneading sequence increases with increasing Λ .

The kneading sequence can be of one of three types

1. It maps to the critical point again, after n iterations. If so, we adopt the convention to terminate the kneading sequence with a C , and refer to the kneading sequence as finite.
2. Preperiodic, i.e., it is infinite but with a periodic tail.
3. Aperiodic.

As an archetype unimodal map we will choose the *tent map*

$$x \mapsto f(x) = \begin{cases} \Lambda x & x \in [0, 1/2] \\ \Lambda(1 - x) & x \in (1/2, 1] \end{cases}, \tag{A18.2}$$

where the parameter $\Lambda \in (1, 2]$. The topological entropy is $h = \log \Lambda$. This follows from the fact any trajectory of the map is bounded, the escape rate is strictly zero, and so the dynamical zeta function

$$1/\zeta(z) = \prod_p \left(1 - \frac{z^{n_p}}{|\Lambda_p|}\right) = \prod_p \left(1 - \left(\frac{z}{\Lambda}\right)^{n_p}\right) = 1/\zeta_{\text{top}}(z/\Lambda)$$

has its leading zero at $z = 1$.

The set of periodic points of the tent map is countable. A consequence of this fact is that the set of parameter values for which the kneading sequence (14.5) is periodic or preperiodic are countable and thus of measure zero and consequently *the kneading sequence is aperiodic for almost all Λ* . For general unimodal maps the corresponding statement is that the kneading sequence is aperiodic for almost all topological entropies.

For a given periodic kneading sequence of period n , $\underline{K}_\Lambda = PC = s_1 s_2 \dots s_{n-1} C$ there is a simple expansion for the topological zeta function. Then the expanded zeta function is a polynomial of degree n

$$1/\zeta_{\text{top}}(z) = \prod_p (1 - z_p^n) = (1 - z) \sum_{i=0}^{n-1} a_i z^i, \quad a_i = \prod_{j=1}^i (-1)^{s_j} \quad (\text{A18.3})$$

and $a_0 = 1$.

Aperiodic and preperiodic kneading sequences are accounted for by simply replacing n by ∞ .

Example. Consider as an example the kneading sequence $K_\Lambda = 10C$. From (A18.3) we get the topological zeta function $1/\zeta_{\text{top}}(z) = (1 - z)(1 - z - z^2)$, see table A18.1. This can also be realized by redefining the alphabet. The only forbidden subsequence is 100. All allowed periodic orbits, except $\bar{0}$, can be built from an alphabet with letters $\underline{10}$ and $\underline{1}$. We write this alphabet as $\{\underline{10}, \underline{1}; \bar{0}\}$, yielding the topological zeta function $1/\zeta_{\text{top}}(z) = (1 - z)(1 - z - z^2)$. The leading zero is the inverse golden mean $z_0 = (\sqrt{5} - 1)/2$.

Example. As another example we consider the preperiodic kneading sequence $K_\Lambda = 101^\infty$. From (A18.3) we get the topological zeta function $1/\zeta_{\text{top}}(z) = (1 - z)(1 - 2z^2)/(1 + z)$, see table A18.1. This can again be realized by redefining the alphabet. There are now an infinite number of forbidden subsequences, namely $101^{2n}0$ where $n \geq 0$. These pruning rules are respected by the alphabet $\{01^{2n+1}; \bar{1}, \bar{0}\}$, yielding the topological zeta function above. The pole in the zeta function $\zeta_{\text{top}}^{-1}(z)$ is a consequence of the infinite alphabet.

An important consequence of (A18.3) is that the sequence $\{a_i\}$ has a periodic tail if and only if the kneading sequence has one (however, their period may differ by a factor of two). We know already that the kneading sequence is aperiodic for almost all Λ .

The analytic structure of the function represented by the infinite series $\sum a_i z^i$ with unity as radius of convergence, depends on whether the tail of $\{a_i\}$ is periodic or not. If the period of the tail is N we can write

$$1/\zeta_{top}(z) = p(z) + q(z)(1 + z^N + z^{2N} \dots) = p(z) + \frac{q(z)}{1 - z^N} ,$$

for some polynomials $p(z)$ and $q(z)$. The result is a set of poles spread out along the unit circle. This applies to the preperiodic case. An aperiodic sequence of coefficients would formally correspond to infinite N and it is natural to assume that the singularities will fill the unit circle. There is indeed a theorem ensuring that this is the case [3], provided the a_i 's can only take on a finite number of values. The unit circle becomes a *natural boundary*, already apparent in a finite polynomial approximations to the topological zeta function, as in figure 18.1. A function with a natural boundary lacks an analytic continuation outside it.

To conclude: The topological zeta function $1/\zeta_{top}$ for unimodal maps has the unit circle as a natural boundary for almost all topological entropies and for the tent map (A18.2), for almost all Λ .

Let us now focus on the relation between the analytic structure of the topological zeta function and the number of periodic orbits, or rather (18.6), the number N_n of fixed points of $f^n(x)$. The trace formula is (see sect. 18.4)

$$N_n = \text{tr } T^n = \frac{1}{2\pi i} \oint_{\gamma_r} dz z^{-n} \frac{d}{dz} \log \zeta_{top}^{-1}$$

where γ_r is a (circular) contour encircling the origin $z = 0$ in clockwise direction. Residue calculus turns this into a sum over zeros z_0 and poles z_p of ζ_{top}^{-1}

$$N_n = \sum_{z_0: r < |z_0| < R} z_0^{-n} - \sum_{z_p: r < |z_p| < R} z_p^{-n} + \frac{1}{2\pi i} \oint_{\gamma_R} dz z^{-n} \frac{d}{dz} \log \zeta_{top}^{-1}$$

and a contribution from a large circle γ_R . For meromorphic topological zeta functions one may let $R \rightarrow \infty$ with vanishing contribution from γ_R , and N_n will be a sum of exponentials.

The leading zero is associated with the topological entropy, as discussed in chapter 18.

We have also seen that for preperiodic kneading there will be poles on the unit circle.

To appreciate the role of natural boundaries we will consider a (very) special example. Cascades of period doublings is a central concept for the description of unimodal maps. This motivates a close study of the function

$$\Xi(z) = \prod_{n=0}^{\infty} (1 - z^{2^n}) . \quad (\text{A18.4})$$

This function will appear again when we derive (A18.3).

The expansion of $\Xi(z)$ begins as $\Xi(z) = 1 - z - z^2 + z^3 - z^4 + z^5 \dots$. The radius of convergence is obviously unity. The simple rule governing the expansion will effectively prohibit any periodicity among the coefficients making the unit circle a natural boundary.

It is easy to see that $\Xi(z) = 0$ if $z = \exp(2\pi m/2^n)$ for any integer m and n . (Strictly speaking we mean that $\Xi(z) \rightarrow 0$ when $z \rightarrow \exp(2\pi m/2^n)$ from inside). Consequently, zeros are dense on the unit circle. One can also show that singular points are dense on the unit circle, for instance $|\Xi(z)| \rightarrow \infty$ when $z \rightarrow \exp(2\pi m/3^n)$ for any integer m and n .

As an example, the topological zeta function at the accumulation point of the first Feigenbaum cascade is $\zeta_{top}^{-1}(z) = (1 - z)\Xi(z)$. Then $N_n = 2^{l+1}$ if $n = 2^l$, otherwise $N_n = 0$. The growth rate in the number of cycles is anything but exponential. It is clear that N_n cannot be a sum of exponentials, the contour γ_R cannot be pushed away to infinity, R is restricted to $R \leq 1$ and N_n is entirely determined by $\int_{\gamma_R}$ which picks up its contribution from the natural boundary.

We have so far studied the analytic structure for some special cases and we know that the unit circle is a natural boundary for almost all Λ . But how does it look out there in the complex plane for some typical parameter values? To explore that we will imagine a journey from the origin $z = 0$ out towards the unit circle. While traveling we let the parameter Λ change slowly. The trip will have a distinct science fiction flavor. The first zero we encounter is the one connected to the topological entropy. Obviously it moves smoothly and slowly. When we move outward to the unit circle we encounter zeros in increasing densities. The closer to the unit circle they are, the wilder and stranger they move. They move from and back to the horizon, where they are created and destroyed through bizarre bifurcations. For some special values of the parameter the unit circle suddenly gets transparent and we get (infinitely) short glimpses of another world beyond the horizon.

We end this section by deriving eqs (A18.5) and (A18.6). The impenetrable prose is hopefully explained by the accompanying tables.

We know one thing from chapter 14, namely for that finite kneading sequence of length n the topological polynomial is of degree n . The graph contains a node which is connected to itself only via the symbol 0. This implies that a factor $(1 - z)$ may be factored out and $\zeta_{top}(z) = (1 - z) \sum_{i=0}^{n-1} a_i z^i$. The problem is to find the coefficients a_i .

The ordered list of (finite) kneading sequences table A18.1 and the ordered list of periodic orbits (on maximal form) are intimately related. In table A18.2 we indicate how they are nested during a period doubling cascade. Every finite kneading sequence PC is bracketed by two periodic orbits, $P1$ and $P0$. We have $\overline{P1} < PC < \overline{P0}$ if P contains an odd number of 1's, and $\overline{P0} < PC < \overline{P1}$ otherwise. From now on we will assume that P contains an odd number of 1's. The other case can be worked out in complete analogy. The first and second harmonic of PC are displayed in table A18.2. The periodic orbit $P1$ (and the corresponding infinite

periodic orbits	finite kneading sequences
$\overline{P1} = A^\infty(P)$	PC
$\overline{P0}$	$P0PC$
$\overline{P0P1}$	$P0P1P0PC$
$\downarrow$ $H^\infty(P)$	$\downarrow$ $H^\infty(P)$

Table A18.2: Relation between periodic orbits and finite kneading sequences in a harmonic cascade. The string P is assumed to contain an odd number of 1's.

$I(C)$	$\zeta_{top}^{-1}(z)/(1-z)$
$P_1 = 100C$	$1 - z - z^2 - z^3$
$H^\infty(P_1) = 10001001100 \dots$	$1 - z - z^2 - z^3 - z^4 + z^5 + z^6 + z^7 - z^8 \dots$
$P' = 10001C$	$1 - z - z^2 - z^3 - z^4 + z^5$
$A^\infty(P_2) = 1000110001 \dots$	$1 - z - z^2 - z^3 - z^4 + z^5 - z^6 - z^7 - z^8 \dots$
$P_2 = 1000C$	$1 - z - z^2 - z^3 - z^4$

Table A18.3: Example of a step in the iterative construction of the list of kneading sequences PC .

kneading sequence) is sometimes referred to as the antiharmonic extension of PC (denoted $A^\infty(P)$) and the accumulation point of the cascade is called the harmonic extension of PC [4] (denoted $H^\infty(P)$).

A central result is the fact that a period doubling cascade of PC is not interfered by any other sequence. Another way to express this is that a kneading sequence PC and its harmonic are adjacent in the list of kneading sequences to any order.

Table A18.3 illustrates another central result in the combinatorics of kneading sequences. We suppose that P_1C and P_2C are neighbors in the list of order 5 (meaning that the shortest finite kneading sequence $P'C$ between P_1C and P_2C is longer than 5.) The important result is that P' (of length $n' = 6$) has to coincide with the first $n' - 1$ letters of both $H^\infty(P_1)$ and $A^\infty(P_2)$. This is exemplified in the left column of table A18.3. This fact makes it possible to generate the list of kneading sequences in an iterative way.

The zeta function at the accumulation point $H^\infty(P_1)$ is

$$\zeta_{P_1}^{-1}(z)\Xi(z^{n_1}) , \quad (\text{A18.5})$$

and just before $A^\infty(P_2)$

$$\zeta_{P_2}^{-1}(z)/(1 - z^{n_2}) . \quad (\text{A18.6})$$

A short calculation shows that this is exactly what one would obtain by applying (A18.3) to the antiharmonic and harmonic extensions directly, provided that it applies to $\zeta_{P_1}^{-1}(z)$ and $\zeta_{P_2}^{-1}(z)$. This is the key observation.

Recall now the product representation of the zeta function $\zeta^{-1} = \prod_p (1 - z^{n_p})$. We will now make use of the fact that the zeta function associated with $P'C$ is a polynomial of order n' . There is no periodic orbit of period shorter than $n' + 1$ between $H^\infty(P_1)$ and $A^\infty(P_2)$. It thus follows that the coefficients of this polynomial coincides with those of (A18.5) and (A18.6), see Table A18.3. We can thus conclude that our rule can be applied directly to $P'C$.

This can be used as an induction step in proving that the rule can be applied to every finite and infinite kneading sequences.

A18.2 Prime factorization for dynamical itineraries



The Möbius function is not only a number-theoretic function, but can be used to manipulate ordered sets of noncommuting objects such as symbol strings. Let $P = \{p_1, p_2, p_3, \dots\}$ be an ordered set of *prime* strings, and

$$\mathcal{N} = \{n\} = \left\{ p_1^{k_1} p_2^{k_2} p_3^{k_3} \cdots p_j^{k_j} \right\},$$

$j \in \mathbb{N}$, $k_i \in \mathbb{Z}_+$, be the set of all strings n obtained by the ordered concatenation of the “primes” p_i . By construction, every string n has a unique prime factorization. We say that a string has a divisor d if it contains d as a substring, and define the string division n/d as n with the substring d deleted. Now we can do things like this: defining $t_n := t_{p_1}^{k_1} t_{p_2}^{k_2} \cdots t_{p_j}^{k_j}$ we can write the inverse dynamical zeta function (23.3) as

$$\prod_p (1 - t_p) = \sum_n \mu(n) t_n,$$

and, if we care (we do in the case of the Riemann zeta function), the dynamical zeta function as

$$\prod_p \frac{1}{1 - t_p} = \sum_n t_n \quad (\text{A18.7})$$

A striking aspect of this formula is its resemblance to the factorization of natural numbers into primes: the relation of the cycle expansion (A18.7) to the product over prime cycles is analogous to the Riemann zeta (exercise 22.7) represented as a sum over natural numbers vs. its Euler product representation.

We now implement this factorization explicitly by decomposing recursively binary strings into ordered concatenations of prime strings. There are 2 strings of length 1, both prime: $p_1 = 0$, $p_2 = 1$. There are 4 strings of length 2: 00, 01, 11, 10. The first three are ordered concatenations of primes: $00 = p_1^2$, $01 = p_1 p_2$, $11 = p_2^2$; by ordered concatenations we mean that $p_1 p_2$ is legal, but $p_2 p_1$ is not. The remaining string is the only prime of length 2, $p_3 = 10$. Proceeding by discarding the strings which are concatenations of shorter primes $p_1^{k_1} p_2^{k_2} \cdots p_j^{k_j}$,

factors	string	factors	string	factors	string	factors	string
p_1	0	p_1^4	0000	p_1^5	00000	$p_1^2 p_5$	00101
p_2	1	$p_1^3 p_2$	0001	$p_1^4 p_2$	00001	$p_1 p_2 p_5$	01101
		$p_1^2 p_2^2$	0011	$p_1^3 p_2^2$	00011	$p_2^2 p_5$	11101
p_1^2	00	$p_1 p_2^3$	0111	$p_1^2 p_2^3$	00111	$p_3 p_5$	10101
$p_1 p_2$	01	p_2^4	1111	$p_1 p_2^4$	01111	$p_1 p_6$	01000
p_2^2	11	$p_1^2 p_3$	0010	p_2^5	11111	$p_2 p_6$	11000
p_3	10	$p_1 p_2 p_3$	0110	$p_1^3 p_3$	00010	$p_1 p_7$	01001
		$p_2^2 p_3$	1110	$p_1^2 p_2 p_3$	00110	$p_2 p_7$	11001
p_1^3	000	p_2^3	1010	$p_1 p_2^2 p_3$	01110	$p_1 p_8$	01011
$p_1^2 p_2$	001	$p_1 p_4$	0100	$p_2^3 p_3$	11110	$p_2 p_8$	11011
$p_1 p_2^2$	011	$p_2 p_4$	1100	$p_1 p_2^2$	01010	p_9	10000
p_2^3	111	$p_1 p_5$	0101	$p_2 p_2^2$	11010	p_{10}	10001
$p_1 p_3$	010	$p_2 p_5$	1101	$p_1^2 p_4$	00100	p_{11}	10010
$p_2 p_3$	110	p_6	1000	$p_1 p_2 p_4$	01100	p_{12}	10011
p_4	100	p_7	1001	$p_2^2 p_4$	11100	p_{13}	10110
p_5	101	p_8	1011	$p_3 p_4$	10100	p_{14}	10111

Table A18.4: Factorization of all periodic points strings up to length 5 into ordered concatenations $p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n}$ of prime strings $p_1 = 0$, $p_2 = 1$, $p_3 = 10$, $p_4 = 100$, $\dots$, $p_{14} = 10111$.

with primes lexically ordered, we generate the standard list of primes, in agreement with table 18.1: 0, 1, 10, 101, 100, 1000, 1001, 1011, 10000, 10001, 10010, 10011, 10110, 10111, 100000, 100001, 100010, 100011, 100110, 100111, 101100, 101110, 101111, $\dots$. This factorization is illustrated in table A18.4.

A18.2.1 Prime factorization for spectral determinants



Following sect. A18.2, the spectral determinant cycle expansions is obtained by expanding F as a multinomial in prime cycle weights t_p

$$F = \prod_p \sum_{k=0}^{\infty} C_{p^k} t_p^k = \sum_{k_1 k_2 k_3 \dots = 0}^{\infty} \tau_{p_1^{k_1} p_2^{k_2} p_3^{k_3} \dots} \quad (\text{A18.8})$$

where the sum goes over all pseudo-cycles. In the above we have defined

$$\tau_{p_1^{k_1} p_2^{k_2} p_3^{k_3} \dots} = \prod_{i=1}^{\infty} C_{p_i^{k_i}} t_{p_i}^{k_i}. \quad (\text{A18.9})$$

exercise 22.7

A striking aspect of the spectral determinant cycle expansion is its resemblance to the factorization of natural numbers into primes: as we already noted in sect. A18.2, the relation of the cycle expansion (A18.8) to the product formula (22.8) is analogous to the Riemann zeta represented as a sum over natural numbers vs. its Euler product representation.

This is somewhat unexpected, as the cycle weights factorize exactly with respect to r repetitions of a prime cycle, $t_{pp\dots p} = t_p^r$, but only approximately (*shadowing*) with respect to subdividing a string into prime substrings, $t_{p_1 p_2} \approx t_{p_1} t_{p_2}$.

The coefficients C_{p^k} have a simple form only in one dimension, given by the Euler formula (28.24). In higher dimensions C_{p^k} can be evaluated by expanding (22.8), $F(z) = \prod_p F_p$, where

$$F_p = 1 - \left(\sum_{r=1}^{\infty} \frac{t_p^r}{r d_{p,r}} \right) + \frac{1}{2} \left(\sum_{r=1}^{\infty} \frac{t_p^r}{r d_{p,r}} \right)^2 - \dots$$

Expanding and recollecting terms, and suppressing the p cycle label for the moment, we obtain

$$\begin{aligned} F_p &= \sum_{r=1}^{\infty} C_k t^k, \quad C_k = (-)^k c_k / D_k, \\ D_k &= \prod_{r=1}^k d_r = \prod_{a=1}^d \prod_{r=1}^k (1 - u_a^r) \end{aligned} \quad (\text{A18.10})$$

where evaluation of c_k requires a certain amount of not too luminous algebra:

$$\begin{aligned} c_0 &= 1 \\ c_1 &= 1 \\ c_2 &= \frac{1}{2} \left(\frac{d_2}{d_1} - d_1 \right) = \frac{1}{2} \left(\prod_{a=1}^d (1 + u_a) - \prod_{a=1}^d (1 - u_a) \right) \\ c_3 &= \frac{1}{3!} \left(\frac{d_2 d_3}{d_1^2} + 2d_1 d_2 - 3d_3 \right) \\ &= \frac{1}{6} \left(\prod_{a=1}^d (1 + 2u_a + 2u_a^2 + u_a^3) \right. \\ &\quad \left. + 2 \prod_{a=1}^d (1 - u_a - u_a^2 + u_a^3) - 3 \prod_{a=1}^d (1 - u_a^3) \right) \end{aligned}$$

etc.. For example, for a general 2-dimensional map we have

$$F_p = 1 - \frac{1}{D_1} t + \frac{u_1 + u_2}{D_2} t^2 - \frac{u_1 u_2 (1 + u_1)(1 + u_2) + u_1^3 + u_2^3}{D_3} t^3 + \dots \quad (\text{A18.11})$$

With $\tau_{p_1^{k_1} p_2^{k_2} \dots p_n^{k_n}}$ defined as above, the prime factorization of symbol strings is unique in the sense that, given a prime ordering convention, *each symbol string* can be written as a unique *concatenation of prime strings*. This factorization is an example of the utility of generalized Möbius inversion, sect. A18.2.

How is the factorization of sect. A18.2 used in practice? Suppose we have computed (or perhaps even measured in an experiment) all prime cycles up to

period n , i.e., we have a list of t_p 's and the corresponding Jacobian matrix eigenvalues $\Lambda_{p,1}, \Lambda_{p,2}, \dots, \Lambda_{p,d}$. A cycle expansion of the Selberg product is obtained by generating all strings in order of increasing length j allowed by the symbolic dynamics and constructing the multinomial

$$F = \sum_n \tau_n \quad (\text{A18.12})$$

where $n = s_1 s_2 \dots s_j$, s_i range over the alphabet, in the present case $\{0, 1\}$. Factorizing every string $n = s_1 s_2 \dots s_j = p_1^{k_1} p_2^{k_2} \dots p_j^{k_j}$ as in table A18.4, and substituting $\tau_{p_1^{k_1} p_2^{k_2} \dots}$ we obtain a multinomial approximation to F . For example, $\tau_{001001010101} = \tau_{001001010101} = \tau_{001^2 \tau_{01^3}}$, and τ_{01^3} , τ_{001^2} are known functions of the corresponding cycle eigenvalues. The zeros of F can now be easily determined by standard numerical methods. The fact that as far as the symbolic dynamics is concerned, the cycle expansion of a Selberg product is simply an average over all symbolic strings makes Selberg products rather pretty.

To be more explicit, we illustrate the above by expressing binary strings as concatenations of prime factors. We start by computing N_n , the number of terms in the expansion (A18.8) of the total cycle period n . Setting $C_{p^k} t_p^k = z^{n_{p^k}}$ in (A18.8), we obtain

$$\sum_{n=0}^{\infty} N_n z^n = \prod_p \sum_{k=0}^{\infty} z^{n_{p^k}} = \frac{1}{\prod_p (1 - z^{n_p})}.$$

So the generating function for the number of terms in the Selberg product is the topological zeta function. For the complete binary dynamics we have $N_n = 2^n$ contributing terms of length n :

$$\zeta_{top} = \frac{1}{\prod_p (1 - z^{n_p})} = \frac{1}{1 - 2z} = \sum_{n=0}^{\infty} 2^n z^n$$

Hence the number of distinct terms in the expansion (A18.8) is the same as the number of binary strings, and conversely, the set of binary strings of length n suffices to label all terms of the total cycle period n in the expansion (A18.8).

A18.3 Counting curvatures

ONE CONSEQUENCE of the finiteness of topological polynomials is that the contributions to curvatures at every order are even in number, half with positive and half with negative sign. For instance, for complete binary labeling (23.8),

$$\begin{aligned} c_4 = & -t_{0001} - t_{0011} - t_{0111} - t_0 t_{01} t_1 \\ & + t_0 t_{001} + t_0 t_{011} + t_{001} t_1 + t_{011} t_1. \end{aligned} \quad (\text{A18.13})$$



We see that 2^3 terms contribute to c_4 , and exactly half of them appear with a negative sign - hence if all binary strings are admissible, this term vanishes in the counting expression.

exercise 23.7

Such counting rules arise from the identity

$$\prod_p (1 + t_p) = \prod_p \frac{1 - t_p^2}{1 - t_p}. \quad (\text{A18.14})$$

Substituting $t_p = z^{n_p}$ and using (18.14) we obtain for unrestricted symbol dynamics with N letters

$$\prod_p (1 + z^{n_p}) = \frac{1 - Nz^2}{1 - Nz} = 1 + Nz + \sum_{k=2}^{\infty} z^k (N^k - N^{k-1})$$

The z^n coefficient in the above expansion is the number of terms contributing to c_n curvature, so we find that for a complete symbolic dynamics of N symbols and $n > 1$, the number of terms contributing to c_n is $(N - 1)N^{n-1}$ (of which half carry a minus sign).

exercise A18.3

We find that for complete symbolic dynamics of N symbols and $n > 1$, the number of terms contributing to c_n is $(N - 1)N^{n-1}$. So, superficially, not much is gained by going from periodic orbits trace sums which get N^n contributions of n to the curvature expansions with $N^n(1 - 1/N)$. However, the point is not the number of the terms, but the cancelations between them.

Commentary

Remark A18.1. Proving the kneading sequence – topological zeta function relations. The explicit relation between the kneading sequence and the coefficients of the topological zeta function is not commonly seen in the literature. The result can be proven by combining some theorems of Milnor and Thurston [5]. That approach is hardly instructive in the context of sect. A18.1. Our derivation was inspired by Metropolis, Stein and Stein classical paper [4]. For further details, consult ref. [1]. (P. Dahlqvist)

Remark A18.2. The XXX inversion formula. One gray day in 1990, in the Bristol University library a graduate student whose name is Jon Keating found an interesting *Physical Review Letter* by Dr. XXX (the policy of ChaosBook.org is not to up citation counts for plagiarized or wrong papers, hence ‘XXX’), entitled “XXX Möbius XXX in physics.” The article starts with the Theorem 268 of Hardy and Wright [2], then derives the generalized inversion formula. By stroke of luck Keating owned the same edition of Hardy and Wright; the generalized inversion formula turned out to be precisely the Theorem 269, the page overleaf. By the evening Keating penned and faxed off a comment to *Phys. Rev. Letters*. The editors response was that *Phys. Rev. Letters* “does not publish comments that are mere factual corrections,” and got instead Dr. XXX to publish an erratum saying that “Equation (7) in the text is equivalent to Theorem 270 in Hardy and Wright [...]. If one starts from this theorem, instead of from the original Möbius theorem, the paper becomes more concise.”

Two weeks later Sir John Maddox, Nature editor, wrote an entire page editorial on the XXX inversion formula, and how marvelous it was that a physicist discovered all this new mathematics. Six months later, *Physical Review A* published a Rapid Communication entitled “On XXX’s inversion formula” by a group from New Zealand. Keating requested *Physical Review* to equip all its referees with a copy of Hardy and Wright, but the proposal was turned down, and ever since there has been a stream of papers on the subject; as of 2017, the paper had over 170 citations.

References

- [1] P. Dahlqvist, “On the effect of pruning on the singularity structure of zeta functions”, *J. Math. Phys.* **38**, 4273–4282 (1997).
- [2] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers* (Oxford Univ. Press, Oxford, 1979).
- [3] E. Hille, *Analytic Function Theory*, Vol. 2 (AMS Chelsea Publishing, 2005).
- [4] N. Metropolis, M. L. Stein, and P. R. Stein, “On finite limit sets for transformations on the unit interval”, *J. Combin. Theory* **15**, 25–44 (1973).
- [5] J. Milnor and W. Thurston, “On iterated maps of the interval”, in *Dynamical Systems*, edited by J. C. Alexander (Springer, New York, 1988), pp. 465–563.

A18.4 Exercises

A18.1. **Lefschetz zeta function.** Elucidate the relation between the topological zeta function and the Lefschetz zeta function.

A18.2. **Cycle expansion denominators.**  Prove that the denominator of c_k is indeed D_k , as asserted (A18.10).

A18.3. **Counting subsets of cycles.** The techniques developed above can be generalized to counting subsets of cycles. Consider the simplest example of a dynamical system with a complete binary tree, a repeller map (14.21) with two straight branches, which we label 0 and 1. Every cycle weight for such map factorizes, with a factor t_0 for each 0, and factor t_1 for each 1 in its symbol string. The transition matrix traces (18.29) collapse to $\text{tr}(T^k) = (t_0 + t_1)^k$, and $1/\zeta$ is simply

$$\prod_p (1 - t_p) = 1 - t_0 - t_1 \quad (\text{A18.15})$$

Substituting into the identity

$$\prod_p (1 + t_p) = \prod_p \frac{1 - t_p^2}{1 - t_p}$$

we obtain

$$\begin{aligned} \prod_p (1 + t_p) &= \frac{1 - t_0^2 - t_1^2}{1 - t_0 - t_1} = 1 + t_0 + t_1 + \frac{2t_0t_1}{1 - t_0 - t_1} \\ &= 1 + t_0 + t_1 + \sum_{n=2}^{\infty} \sum_{k=1}^{n-1} 2 \binom{n-2}{k-1} t_0^k t_1^{n-k}. \end{aligned} \quad (\text{A18.16})$$

Hence for $n \geq 2$ the number of terms in the expansion ?? with k 0's and $n - k$ 1's in their symbol sequences is $2 \binom{n-2}{k-1}$. This is the degeneracy of distinct cycle eigenvalues in fig.?!; for systems with non-uniform hyperbolicity this degeneracy is lifted (see fig. ?!).

In order to count the number of prime cycles in each such subset we denote with $M_{n,k}$ ($n = 1, 2, \dots$; $k = \{0, 1\}$ for $n = 1$; $k = 1, \dots, n - 1$ for $n \geq 2$) the number of prime n -cycles whose labels contain k zeros, use binomial string counting and Möbius inversion and obtain

$$\begin{aligned} M_{1,0} &= M_{1,1} = 1 \\ nM_{n,k} &= \sum_{m \mid \frac{n}{k}} \mu(m) \binom{n/m}{k/m}, \quad n \geq 2, k = 1, \dots, n - 1 \end{aligned}$$

where the sum is over all m which divide both n and k .

Appendix A20

Averaging

HERE WE REVIEW some elementary notions of probability theory that will be useful to you no matter what you do with the rest of your life.

A20.1 Moments, cumulants



The exact *mean* μ (or expectation or expected value $\mathbb{E}[a]$) is the integral of the random variable a with respect to its probability measure ρ (20.7), commonly denoted

$$\mu = \mathbb{E}[a] = \langle a \rangle = \int_{\mathcal{M}} dx \rho(x) a(x). \quad (\text{A20.1})$$

In deterministic chaos nothing is random: $a(x)$ is what we call ‘observable’, introduced in sect. 20.1.1, and $\rho(x)$ is the deterministic weight explained in chapter 23. We use $\langle \cdots \rangle_{\rho}$ or simply $\langle \cdots \rangle$ to denote an integral over state space weighted by ρ , while overbar $\overline{\cdots}$ denotes a time average. If the average is over a (finite or infinite) set of states labeled by discrete labels π , each state contributing with a weight t_{π} , the expectation is given by

$$\langle a \rangle = \sum_{\pi} a_{\pi} t_{\pi}, \quad (\text{A20.2})$$

with probabilities in either case normalized so that $\langle 1 \rangle = 1$.

The expectation $\langle a^k \rangle$ is called the k th *moment*. The first moment is the mean μ defined in (A20.1). For $k > 1$, it is more natural to consider the moments about the mean, $\langle (a - \langle a \rangle)^k \rangle$, called *central moments*. The second, and all-important central moment is known as the *variance*,

$$\sigma^2 = \langle (a - \langle a \rangle)^2 \rangle = \langle a^2 \rangle - \langle a \rangle^2, \quad (\text{A20.3})$$

or, in probabilist notation,

$$\mathbb{E}[a^2] = \mu^2 + \sigma^2. \quad (\text{A20.4})$$

Its positive square root σ is called the *standard deviation* σ . As a mnemonic, think of the width of a Gaussian being $\approx 2\sigma$.

Standardized moment

$$\langle (a - \langle a \rangle)^k \rangle / \sigma^k \quad (\text{A20.5})$$

is the k th central moment divided by σ^k , a dimensionless representation of the distribution of variance 1, independent of translations and linear changes of scale.

Moments can be collected into the (exponential) moment-generating function

$$\langle e^{\beta a} \rangle = 1 + \sum_{k=1}^{\infty} \frac{\beta^k}{k!} \langle a^k \rangle. \quad (\text{A20.6})$$

question A20.1

Why the prefactor $1/k!$ (a Taylor series), and not $1/k$ (a logarithmic series), or 1 (discrete Laplace transform or Z-transform)? In statistical, stochastic and quantum mechanics / quantum field theory applications one is solving linear ODEs or PDEs, and their solutions are always exponential in form.

Hardly any experiment measures a^k for $k > 2$ -that might require a lot of data- and raising approximate numbers to high powers is not smart: if $|a| < 1$, a^k gets very small very fast, and conversely if $|a| > 1$, a^k gets very big. Still, with a bit of hindsight, one finds that moments do play a natural, fundamental role if folded into the *cumulant-generating function*

remark A20.1

$$\ln \langle e^{\beta a} \rangle = \sum_{k=1}^{\infty} \frac{\beta^k}{k!} \langle a^k \rangle_c, \quad (\text{A20.7})$$

where the subscript c indicates a *cumulant*, or, in statistical mechanics and quantum field theory contexts, the ‘connected Green’s function’. Were $\langle a^k \rangle = \langle a \rangle^k$, we would have only one term in the series (A20.7), $\ln \langle e^{\beta a} \rangle = \ln e^{\beta \langle a \rangle} = \beta \langle a \rangle$, and that would be that. So cumulants $\langle a^k \rangle_c$ measure fluctuations about the mean $\langle a \rangle$. Indeed, expanding the logarithm of the series (A20.6), it is easy to check that the first cumulant is the mean, the second is the variance,



$$\langle a^2 \rangle_c = \langle (a - \langle a \rangle)^2 \rangle = \langle a^2 \rangle - \langle a \rangle^2 = \sigma^2, \quad (\text{A20.8})$$

and $\langle a^3 \rangle_c$ is the third central moment, or the *skewness*,

$$\langle a^3 \rangle_c = \langle (a - \langle a \rangle)^3 \rangle = \langle a^3 \rangle - 3\langle a^2 \rangle \langle a \rangle + 2\langle a \rangle^3. \quad (\text{A20.9})$$

The higher cumulants, however, are *not* central moments. The fourth cumulant,

$$\begin{aligned} \langle a^4 \rangle_c &= \langle (a - \langle a \rangle)^4 \rangle - 3\langle (a - \langle a \rangle)^2 \rangle^2 \\ &= \langle a^4 \rangle - 4\langle a^3 \rangle \langle a \rangle - 3\langle a^2 \rangle^2 + 12\langle a^2 \rangle \langle a \rangle^2 - 6\langle a \rangle^4, \end{aligned} \quad (\text{A20.10})$$

rewritten in terms of standardized moments, is known as the *kurtosis*:

$$\frac{1}{\sigma^4} \langle a^4 \rangle_c = \frac{1}{\sigma^4} \langle (a - \langle a \rangle)^4 \rangle - 3. \quad (\text{A20.11})$$

The deep reason why cumulants are preferable to moments is that for a normalized Gaussian distribution all cumulants beyond the second one vanish, so they are a measure of deviation of statistics from Gaussian (see example 24.3). For a ‘free’ or ‘Gaussian’ field theory the only non-vanishing cumulant is the second one; for field theories with interactions the derivatives of $\ln \langle \exp(\beta a) \rangle$ with respect to β then yield cumulants, or the Burnett coefficients (24.14), or ‘effective’ n -point Green’s functions or n -point correlations.

exercise A20.1

question A20.2

So, what’s so special about Gaussians?



example A20.1
p. 999



example A20.2
p. 999

A20.1.1 Covariance matrix

For a multi-component observable, the second central moment is called the *covariance matrix*

$$Q_{ij} = \langle (a_i - \langle a_i \rangle)(a_j - \langle a_j \rangle) \rangle. \quad (\text{A20.12})$$

As Q is a symmetric, diagonalizable matrix, with eigenvalues σ_k^2 and orthogonal eigenvectors $\mathbf{e}^{(k)}$, you can visualize such *multivariate normal distribution* as a cigar-shaped cloud of points, with orthonormal principal axes of standard deviation (singular value) lengths σ_k . A cigar fat in a few directions, negligibly thin in the remaining directions motivates reduced-dimensional, linear modeling of the data by a retaining only a hyperplane spanned by the dominant directions; depending on the community, this is called the principal component analysis (PCA), the proper orthogonal decomposition (POD), the singular value decomposition (SVD), or the [Karhunen–Loève transform](#).

section 6.1

A20.1.2 Empirical means

Given a set of N iid (independently identically distributed) data samples $\{a_i\}$, where “iid” means that probability measures ρ factorize,

$$\rho(a_i, a_j) = \rho(a_i) \rho(a_j), \quad i \neq j, \quad (\text{A20.13})$$

the *empirical mean* of observable a is the average

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N a_i. \quad (\text{A20.14})$$

example A20.3
p. 1000

$\hat{\mu}$ is *unbiased* if $\mathbb{E}[\hat{\mu}] = \mu$; we verify that in example A20.4. However, the *unbiased sample variance* $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$ of observable a is defined differently, as

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (a_i - \hat{a})^2. \quad (\text{A20.15})$$

What's up with the $N - 1$ divisor? See

example A20.4
p. 1002example A20.5
p. 1003

Commentary

Remark A20.1. Cumulants are ChaosBook's central leitmotif, see sect. 18.3, sect. 23.2.2, remark 23.2, sect. 28.1, sect. 41.2.1, and appendix A40.4. For details and background on cumulants of random variables, consult the *cumulant* [wiki](#), and the *Ursell functions* or *connected correlation functions* [wiki](#).

In field theory cumulants can be expressed as sums over connected Green's functions, or connected Feynman diagrams (see *exponential formula* [wiki](#)).

The power series for the full partition function $\langle Z \rangle$ in terms of the connected $\ln \langle Z \rangle$ partition function is given by [Bell polynomials](#) [2, 5–7]

$$\exp\left(\sum_{m=1}^{\infty} x_m \frac{t^m}{m!}\right) = \sum_{n=0}^{\infty} B_n(x_1, x_2, \dots, x_n) \frac{t^n}{n!}, \quad (\text{A20.16})$$

with n th moment (A20.1) expressed in terms of cumulants $x_n = \langle a^n \rangle_c$ by

$$\langle a^n \rangle = B_n(x_1, x_2, \dots, x_n). \quad (\text{A20.17})$$

Bell polynomial expressions for $\ln \langle Z \rangle$ are, however, seem rather inefficient, and obscure the fact that the cumulant expansion of $\ln \langle Z \rangle$ is simpler than that of $\langle Z \rangle$.

Remark A20.2. Gaussian integrals. Kadanoff [4] has a nice discussion of Gaussian integrals, the central limit theorem and large deviations in *Chap. 3 Gaussian Distributions*, available online [here](#).

Question A20.1. Alice asks

Q Isn't expectation value (A20.6) the *characteristic function*?

A With imaginary exponent, $\beta \rightarrow it$ and the observable defined in the momentum space, $a = a(p)$, the expectation value (A20.6) does have the form of a [characteristic function](#), i.e., the Fourier transform of the probability density function

$$\mathbb{E}[e^{ipx}] = \int_{\mathcal{M}} dx \rho(x) e^{ipx}. \quad (\text{A20.18})$$

Question A20.2. Alice muses

Q Somehow cumulants seem to fill my head with ideas — only I don't exactly know what they are!

A A scholarly aside, safely ignored, on where the characteristic state function $s(\beta)$ (20.10) fits into the grander scheme of things: in statistical mechanics and field theory, the partition function and the Helmholtz free energy have form

$$Z(\beta) = \exp(-\beta F), \quad F(\beta) = -\frac{1}{\beta} \ln Z(E), \quad (\text{A20.19})$$

so in that sense $\langle e^{\beta a} \rangle$ is a ‘partition function’, and $s(\beta)$ in (20.10) is the associated ‘free energy’. Expanding the logarithm of the series (A20.6) is easy for the first few terms, but it quickly gets old. The smart way to do this, explained in ref. [1], is to write down the Dyson-Schwinger equations that generate recursively the terms in the Helmholtz free energy expansion (connected Green’s functions) and Gibbs free energy (1-particle irreducible Green’s functions).

References

- [1] P. Cvitanović, *Field Theory*, Notes prepared by E. Gyldenkerne (Nordita, Copenhagen, 1983).
- [2] B. Fiol, J. Martínez-Montoya, and A. Rios Fukelman, “Wilson loops in terms of color invariants”, *J. High Energy Phys.* **2019**, 202 (2019).
- [3] E. V. Ivashkevich, N. S. Izmailian, and C.-K. Hu, “Kronecker’s double series and exact asymptotic expansions for free models of statistical mechanics on torus”, *J. Phys. A* **35**, 5543–5561 (2002).
- [4] L. P. Kadanoff, *Statistical Physics: Statics, Dynamics and Renormalization* (World Scientific, Singapore, 2000).
- [5] E. K. Lloyd, *Bell polynomial*, Encyclopedia of Mathematics, 2011.
- [6] N. Wheeler, *Bell Polynomials and Related Constructs*, 2020.
- [7] C. S. Withers and S. Nadarajah, “Moments from cumulants and vice versa”, *Internat. J. Math. Ed. Sci. Tech.* **40**, 842–845 (2009).

A20.2 Examples

Example A20.1. Relation between moments and cumulants. Moments Z_k and cumulants Q_k

$$1 + \sum_{k=1}^{\infty} \frac{z^k}{k!} Z_k = \exp \left\{ \sum_{k=1}^{\infty} \frac{z^k}{k!} Q_k \right\} \quad (\text{A20.20})$$

are related to each other as

$$\begin{aligned} Z_1 &= Q_1 \\ Z_2 &= Q_2 + Q_1^2 \\ Z_3 &= Q_3 + 3Q_1Q_2 + Q_1^3 \\ Z_4 &= Q_4 + 4Q_1Q_3 + 3Q_2^2 + 6Q_1^2Q_2 + Q_1^4 \\ &\vdots \\ Z_k &= \sum_{r=1}^k \sum \left(\frac{Q_{k_1}}{k_1!} \right)^{i_1} \cdots \left(\frac{Q_{k_r}}{k_r!} \right)^{i_r} \frac{k!}{i_1! \cdots i_r!}, \end{aligned} \quad (\text{A20.21})$$

where summation is over all positive integers $\{i_1 \dots i_r\}$ and $\{k_1, \dots, k_r\}$ such that $k_1 i_1 + \dots + k_r i_r = k$.

[return to p. 996](#)

(Ivashkevich, Izmailian and Hu [3].)

Example A20.2. Gaussian minimizes information. Shannon [information entropy](#) is given by

$$S[\rho] = -\langle \ln \rho \rangle = - \int_{\mathcal{M}} dx \rho(x) \ln \rho(x), \quad (\text{A20.22})$$

where ρ is a probability density. Shannon thought of $-\ln \rho$ as ‘information’, very roughly in the sense that if -for example- $\rho(x) = 2^{-6}$, it takes $-\ln \rho = 6$ binary bits of ‘information’ to specify the probability density ρ at the point x . Information entropy (A20.22) is the expectation value (or average) of information.

A function $\rho \geq 0$ is an arbitrary function, of which we only require that it is normalized as a probability,

$$\int_{\mathcal{M}} dx \rho(x) = 1, \quad (\text{A20.23})$$

has a mean value,

$$\int_{\mathcal{M}} dx x \rho(x) = \mu, \quad (\text{A20.24})$$

and has a variance

$$\int_{\mathcal{M}} dx x^2 \rho(x) = \mu^2 + \sigma^2. \quad (\text{A20.25})$$

As ρ can be arbitrarily wild, it might take much “information” to describe it. Is there a function $\rho(x)$ that contains the *least* information, i.e., that minimizes the information entropy (A20.22)?

Participant	Stress (X)	Satisfaction (Y)
1	11	7
2	25	1
3	19	4
4	7	9
5	23	2
6	6	8
7	11	8
8	22	3
9	25	3
10	10	6

Table A20.1: Stress (1 to 30 scale) vs. happiness (1 to 10 scale) for a sample of 10 participants.

To find it, we minimize (A20.22) subject to constraints (A20.23)-(A20.25), implemented by adding Lagrange multipliers λ_j

$$\begin{aligned}
 C[\rho] = & \int_{\mathcal{M}} dx \rho(x) \ln \rho(x) \\
 & + \lambda_0 \left(\int_{\mathcal{M}} dx \rho(x) - 1 \right) + \lambda_1 \left(\int_{\mathcal{M}} dx x \rho(x) - \mu \right) \\
 & + \lambda_2 \left(\int_{\mathcal{M}} dx x^2 \rho(x) - \mu^2 - \sigma^2 \right), \quad (\text{A20.26})
 \end{aligned}$$

and looking for the extremum $\delta C = 0$,

$$\frac{\delta C[\rho]}{\delta \rho(x)} = (\ln \rho(x) + 1) + \lambda_0 + \lambda_1 x + \lambda_2 x^2 = 0, \quad (\text{A20.27})$$

so

$$\rho(x) = e^{-(1+\lambda_0+\lambda_1 x+\lambda_2 x^2)}. \quad (\text{A20.28})$$

The Lagrange multipliers λ_j can be expressed in terms of distribution parameters μ and σ by substituting this $\rho(x)$ into the constraint equations (A20.23)-(A20.25). We find that the probability density that minimizes information entropy is the Gaussian

$$\rho(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (\text{A20.29})$$

In what sense is that the distribution with the ‘least information’? As we saw in the derivation of the [cumulant expansion](#) eq. (20.17), for a Gaussian distribution all cumulants but the mean μ and the variance σ^2 vanish, it is a distribution specified by only two ‘informations’, the location of its peak and its width.

[return to p. 996](#)

Sara A. Solla

Example A20.3. I get stress, but I can’t get no satisfaction. A group of participants in a study of the correlation between stress and life satisfaction completed a questionnaire on how stressed they felt, and how satisfied they felt with their lives. Participants’ scores are given in table [A20.1](#).



We start our statistical analysis in the usual way, by evaluating the empirical means (A20.14) of the stress and satisfaction,

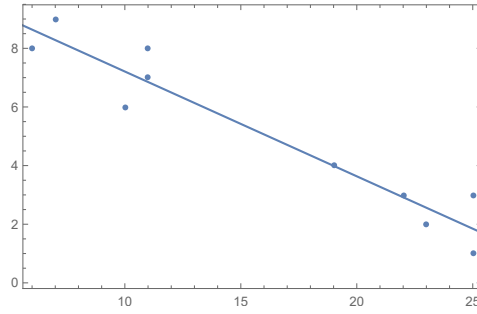
$$\hat{\mu}_X = \frac{1}{10} \sum_{i=1}^{10} X_i = 15.9, \quad \hat{\mu}_Y = \frac{1}{10} \sum_{i=1}^{10} Y_i = 5.1,$$

and the *unbiased* variances (A20.15) and standard deviations,

$$\hat{\sigma}_X^2 = \frac{1}{10-1} \sum_{i=1}^{10} (X_i - \hat{\mu}_X)^2 = 58.1, \quad \hat{\sigma}_X = 7.6.$$

$$\hat{\sigma}_Y^2 = \frac{1}{10-1} \sum_{i=1}^{10} (Y_i - \hat{\mu}_Y)^2 = 8.1, \quad \hat{\sigma}_Y = 2.8.$$

The means are halfway their respective ranges, but the standard deviations are huge, they span across the available ranges. To figure out what is going on, one should always start with *visualizing* the data:



So the empirical means are meaningless - the subjects are either unhappy or happy, there is nobody in between. Standard deviations of such bimodal distributions are not helpful either, as they are measuring deviations from the non-existent average participant. However, the linear fit

$$Y = 11 - .36X \quad (\text{A20.30})$$

is pretty good. Clearly this is 2-dimensional data set, so we compute the stressed/happy covariance

$$Q_{XY} = \frac{1}{10-1} \sum_{i=1}^{10} (X_i - \hat{\mu}_X)(Y_i - \hat{\mu}_Y) = -20.8.$$

Note that the covariance between X and Y is negative. The ellipsoid given by the covariance matrix

$$Q = \begin{pmatrix} \hat{\sigma}_X^2 & Q_{XY} \\ Q_{XY} & \hat{\sigma}_Y^2 \end{pmatrix},$$

with singular values (square roots of eigenvalues) and eigenvectors

$$\{\sigma_1, \sigma_2\} = \{8.10, 0.76\} : \quad \mathbf{e}_{(1)} = (0.94, -0.34), \quad \mathbf{e}_{(2)} = (0.34, 0.94),$$

gives a good description of the data, aligned along $\mathbf{e}_{(1)}$ (of slope close to the linear fit (A20.30)), with small transverse fluctuations along $\mathbf{e}_{(2)}$. The only problem is that we are plotting lemons vs. roses.

For this reason, statisticians like to study pairwise Pearson *correlation coefficients*, such as

$$\rho_{XY} = \frac{Q_{XY}}{\hat{\sigma}_X \hat{\sigma}_Y} = -0.9573,$$

for which the deviation $1 - |\rho_{XY}|$ is a measure for how well the data is fit by a linear fit.

One might be tempted to study the full *correlation coefficients matrix*, a somewhat contrived “standardized” or “whitened” rescaling (A20.5) of the covariance matrix (A20.12),

$$\text{Corr}(X, Y) = \begin{pmatrix} 1 & \rho_{XY} \\ \rho_{XY} & 1 \end{pmatrix} = \begin{pmatrix} \hat{\sigma}_X & 0 \\ 0 & \hat{\sigma}_Y \end{pmatrix}^{-1} \begin{pmatrix} \hat{\sigma}_X^2 & Q_{XY} \\ Q_{XY} & \hat{\sigma}_Y^2 \end{pmatrix} \begin{pmatrix} \hat{\sigma}_X & 0 \\ 0 & \hat{\sigma}_Y \end{pmatrix}^{-1}.$$

Its eigenvalues $\{1 + \rho_{XY}, 1 - \rho_{XY}\}$ and eigenvectors are a *dimensionless* least-squares fit to the data, with the ellipsoid’s principal axes along the diagonals

$$\{\sigma_1, \sigma_2\} = \{1.40, 0.207\} : \quad \mathbf{e}_{(1)} = \frac{1}{\sqrt{2}}(1, -1), \quad \mathbf{e}_{(2)} = \frac{1}{\sqrt{2}}(1, 1).$$

The pairwise *correlation coefficients* have some utility in singling out the signs of slopes in which data is nearly linear ($\rho_{X_i X_j}$ close to ± 1). The transformation from the covariance matrix to the correlations matrix is not a similarity transformation, so while the covariance matrix is a fundamental object in the multivariate cumulant expansions, the correlation matrices are not used in physics, only the 2-dimensional planes spanned by $\rho_{X_i X_j}$ are informative.

Bonus reading: “The Economist” May 2, 2013 article (if you can get past the pay-wall), or, more seriously, D. Kahneman and A. Deaton -the 2002 Nobel Memorial Prize in Economic Sciences- about the correlation between income and happiness.

Citing from the Economist: “THE Easterlin paradox, named for economist Richard Easterlin, reckons that higher incomes do not necessarily make people happier. Since Mr Easterlin first made his conjecture in 1974, economists’ views have evolved: money matters, studies suggest, but only up to a point. Become rich enough, and a bigger paycheque no longer leads to more happiness. Yet a new NBER *working paper* by economists Betsey Stevenson and Justin Wolfers, both of the University of Michigan, casts doubt on this chestnut. They use a trove of data generated by Gallup, a polling firm, from its *World Poll*. Gallup asked respondents around the world to imagine a ‘satisfaction ladder’ in which the top step represents a respondent’s best possible life. Those being polled are then asked where on the ladder they stand (from zero to a maximum of 10), and how much they earn. Though some countries seem happier than others, people everywhere report more satisfaction as they grow richer. Even more striking, the relationship between income and happiness hardly changes as incomes rise. Moving from rich to richer seems to raise happiness just as much as moving from poor to less poor. One never really grows tired of earning more.”

Table A20.1 is at odds with this conclusion. Maybe Ignacio just made it up? Penny for your thoughts.

return to p. 997

I. Taboada, P. Cvitanović & S.A. Solla

Example A20.4. Unbiased sample variance.

Why is the empirical estimate for the *unbiased sample variance*

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (a_i - \hat{\mu})^2 \quad (\text{A20.31})$$

defined with the $N - 1$ divisor?

At this point your instructor mumbled something about “degrees of freedom” and moved on, but why mumble if you can compute? By the definition (A20.1), expectations of unbiased estimates are exact,

$$\mathbb{E}[\hat{\mu}] = \mu, \quad \mathbb{E}[\hat{\sigma}^2] = \sigma^2. \quad (\text{A20.32})$$

That is true of the empirical mean (A20.14),

$$\mathbb{E}[\hat{\mu}] = \frac{1}{N} \sum_{i=1}^N \mathbb{E}[a_i] = \frac{1}{N} \sum_{i=1}^N \mu = \mu,$$

but the empirical estimate for the sample variance written as average over the sum of deviations square does not quite work out. Assume first that the empirical variance is given by the usual average

$$\begin{aligned} \bar{\sigma}^2 &= \frac{1}{N} \sum_{i=1}^N (a_i - \hat{\mu})^2 = \frac{1}{N} \sum_{i=1}^N (a_i^2 - 2\hat{\mu}a_i + \hat{\mu}^2) = \frac{1}{N} \sum_{i=1}^N a_i^2 - \frac{1}{N^2} \left(\sum_{i=1}^N a_i \right)^2 \\ &= \frac{1}{N} \sum_{i=1}^N a_i^2 - \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N a_i a_j = \frac{N-1}{N^2} \sum_{i=1}^N a_i^2 - \frac{1}{N^2} \sum_{i \neq j}^N a_i a_j. \end{aligned} \quad (\text{A20.33})$$

By the iid independence of individual measurements (A20.13), and $\sigma^2 = \mathbb{E}[a^2] - \mu^2$ relation (A20.4), the expectation of $\bar{\sigma}^2$ is

exercise A20.2

$$\begin{aligned} \mathbb{E}[\bar{\sigma}^2] &= \frac{N-1}{N^2} \sum_{i=1}^N \mathbb{E}[a_i^2] - \frac{1}{N^2} \sum_{i \neq j}^N \mathbb{E}[a_i a_j] = \frac{N-1}{N} (\mathbb{E}[a_i^2] - \mathbb{E}[a_i] \mathbb{E}[a_j]) \\ &= \frac{N-1}{N} (\mathbb{E}[a^2] - \mu^2) = \frac{N-1}{N} \sigma^2. \end{aligned} \quad (\text{A20.34})$$

This attempt at a definition of empirical variance $\bar{\sigma}^2$ thus violates the ‘unbiased’ condition (A20.32). The *unbiased empirical variance* (A20.15), $\hat{\sigma}^2 = N\bar{\sigma}^2/(N-1)$, is correct for any sample size, not only in the $N \rightarrow \infty$ limit. What happened? a_i, a_j are iid only for the $N^2 - N$ off-diagonal covariance elements; the squares a_j^2 along the diagonal do not contribute to “covariance.”

return to p. 997

(continued in example A20.5)

Sara A. Solla

Example A20.5. Standard error of the mean. (Continued from example A20.4)

Think now of estimating the empirical mean (A20.14) of observable a as $j = 1, 2, \dots, N$ attempts to estimate the mean $\hat{\mu}_j$, each based on M data samples

$$\hat{\mu}_j = \frac{1}{M} \sum_{i=1}^M a_i. \quad (\text{A20.35})$$

Every attempt yields a different sample mean, so $\hat{\mu}_j$ itself is an iid random variable, with unbiased expectation $\mathbb{E}[\hat{\mu}] = \mu$. What is its variance

$$\text{Var}[\hat{\mu}] = \mathbb{E}[(\hat{\mu} - \mu)^2] = \mathbb{E}[\hat{\mu}^2] - \mu^2 ?$$

This calculation is very much the same as the one carried out in example [A20.4](#), resulting in

$$\text{Var}[\hat{\mu}] = \frac{1}{N} \sigma^2$$

[exercise A20.3](#)

The quantity $\sqrt{\text{Var}[\hat{\mu}]} = \sigma / \sqrt{N}$ is called the *standard error of the mean* (SEM); it tells us that the accuracy of the determination of the mean μ increases as the $1/\sqrt{N}$, where N is the number of estimate attempts, each based on the same number of data points.

[return to p. 997](#)

Sara A. Solla

A20.3 Exercises

A20.1. **Cumulants.** Show that for a Gaussian probability distribution (a) all odd moments vanish, and (b) all cumulants in (A20.7) vanish for $n \geq 3$, $\langle a^n \rangle_c = 0$.

(P. Cvitanović)

A20.2. **Unbiased sample variance.** Empirical estimates of the mean $\hat{\mu}$ and the variance $\hat{\sigma}^2$ are said to be “unbiased” if their expectations equal the exact values,

$$\mathbb{E}[\hat{\mu}] = \mu, \quad \mathbb{E}[\hat{\sigma}^2] = \sigma^2. \quad (\text{A20.36})$$

(a) Verify that the empirical mean

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N a_i \quad (\text{A20.37})$$

is unbiased.

(b) Show that the naive empirical estimate for the *sample variance*

$$\bar{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (a_i - \hat{\mu})^2 = \frac{1}{N} \sum_{i=1}^N a_i^2 - \frac{1}{N^2} \left(\sum_{i=1}^N a_i \right)^2$$

is biased. Hint: note that in evaluating $\mathbb{E}[\dots]$ you have to separate out the diagonal terms in

$$\left(\sum_{i=1}^N a_i \right)^2 = \sum_{i=1}^N a_i^2 + \sum_{i \neq j} a_i a_j. \quad (\text{A20.38})$$

(c) Show that the empirical estimate of form

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (a_i - \hat{\mu})^2, \quad (\text{A20.39})$$

is unbiased.

(d) Is this empirical sample variance unbiased for any finite sample size, or is it unbiased only in the $N \rightarrow \infty$ limit?

Sara A. Solla

A20.3. **Standard error of the mean.**

Now, estimate the empirical mean (A20.37) of observable a by $j = 1, 2, \dots, N$ attempts to estimate the mean $\hat{\mu}_j$, each based on M data samples

$$\hat{\mu}_j = \frac{1}{M} \sum_{i=1}^M a_i. \quad (\text{A20.40})$$

Every attempt yields a different sample mean.

(a) Argue that $\hat{\mu}_j$ itself is an iid random variable, with unbiased expectation $\mathbb{E}[\hat{\mu}] = \mu$.

(b) What is its variance

$$\text{Var}[\hat{\mu}] = \mathbb{E}[(\hat{\mu} - \mu)^2] = \mathbb{E}[\hat{\mu}^2] - \mu^2$$

as a function of variance expectation (A20.36) and N , the number of $\hat{\mu}_j$ estimates? Hint; one way to do this is to repeat the calculations of exercise A20.2, this time for $\hat{\mu}_j$ rather than a_i .

(c) The quantity $\sqrt{\text{Var}[\hat{\mu}]} = \sigma / \sqrt{N}$ is called the *standard error of the mean* (SEM); it tells us about the accuracy of the determination of the mean μ . How does the SEM decrease as the N , the number of estimate attempts, increases?

Sara A. Solla

Appendix A22

Spectral determinants

WE REDERIVE HERE the dynamical zeta function (22.11) by the *transfer operator* technique. This is done only to accommodate the reader versed in statistical mechanics transfer operators; this appendix can be skipped in good conscience by anybody else.

A22.1 Transfer operators



Consider Mandelbrot's favorite example of a fractal, the Cantor set. The set is generated by a single rule: replace a “mother” interval l by two “daughters” of length $l/3$; repeat this replacement *ad infinitum*. Given the rule, one can immediately compute the Hausdorff dimension; at n th level the set can be covered with 2^n intervals of size 3^{-n} , hence $D = \log 2 / \log 3$.

A transfer operator is a generalization of such rule to non-wandering sets for which the dynamics generates an infinity of scales, not just a single scale as in the Cantor set case. For example, for a repeller like the one illustrated in figure 14.9 the dynamics associates with each “mother” interval $\mathcal{M}_m$, $m = s_2 s_3 \cdots s_n$, two “daughter” intervals $\mathcal{M}_d$, $d = 0 s_2 \cdots s_n$, $1 s_2 \cdots s_n$, at the next level of resolution. The transfer operator appropriate to the evaluation of (27.2) is defined by the set of daughter/mother ratios

$$T_{dm} = \mathcal{M}_d / \mathcal{M}_m \quad (\text{A22.1})$$

For the Cantor set $T_{dm} = 1/3$ for all d ; for a generic dynamical non-wandering set T_{dm} takes on infinity of values. The sum (27.2) can now be expressed in terms of products of transfer operators:

$$\Gamma_n = \sum_{s_1 s_2 \cdots s_n} T_{s_1 s_2 \cdots s_n, s_2 \cdots s_n} T_{s_2 \cdots s_n, s_3 \cdots s_n} \cdots T_{s_n, \cdot} \quad (\text{A22.2})$$

As it stands, this is a purely formal rewrite of (27.2); the “mother” to “daughters” relations place the pieces of a non-wandering set onto a hierarchical tree, and that can be done in various ways. To proceed, we require that the tree provide a hierarchical nesting of the scaling ratios in the following sense: the value of $T_{s_1 s_2 \dots s_n, s_2 \dots s_n}$ should depend strongly on the head of the symbol sequence $s_1 s_2 \dots$, and weakly on the tail $\dots s_{n-1} s_n$. More precisely, we assume that the specification of first k symbols determines $T_{s_1 s_2 \dots s_n, s_2 \dots s_n}$, $n > k$, within accuracy Δ_k

$$T_{s_1 s_2 \dots s_n, s_2 \dots s_n} = \tilde{T}_{s_1 s_2 \dots s_k, s_2 \dots s_{k+1}}^{(k)} + O(\Delta_k), \quad (\text{A22.3})$$

and that $|\Delta_k|$ decrease monotonically towards zero with increasing k . Here $\tilde{T}^{(k)}$ is an approximate “mean” scaling for all T_{dm} with the same first k symbols. Replacing the infinite number of scaling ratios (A22.1) by a finite matrix $\tilde{T}^{(k)}$ amounts to approximating the non-wandering set by a Cantor set with a finite number of scales.

An example of such hierarchy is the 1-dimensional repeller of figure 14.9, for which $T_{dm} \approx 1/|f'(x)|$, where $f'(x)$ is a slope of the mapping evaluated at a point x inside the d th neighborhood. With the labeling conventions of figure 14.9, the points whose itineraries have the same head $s_1 s_2 \dots s_n$ are spatially close, and hence the associated derivatives and transfer matrix elements are close.

Now we can study the transfer operator T as a limit of $\tilde{T}^{(k)}$ finite matrix approximations. For example, for the binary labeled repeller of figure 14.9, $k = 2$ level approximation to T is given by

$$\tilde{T}^{(2)} = \begin{bmatrix} \tilde{T}_{00.00} & \tilde{T}_{00.01} & 0 & 0 \\ 0 & 0 & \tilde{T}_{01.10} & \tilde{T}_{01.11} \\ \tilde{T}_{10.00} & \tilde{T}_{10.01} & 0 & 0 \\ 0 & 0 & \tilde{T}_{11.10} & \tilde{T}_{11.11} \end{bmatrix}$$

For binary symbolic dynamics $\tilde{T}$ is in general a sparse matrix, as the only nonvanishing entries in the $m = s_2 s_3 \dots s_{k+1}$ column of $\tilde{T}_{dm}$ are in the rows $0 s_2 \dots s_k$ and $1 s_2 \dots s_k$.

In the k th order approximation the sums in (A22.2) reduce to matrix multiplication

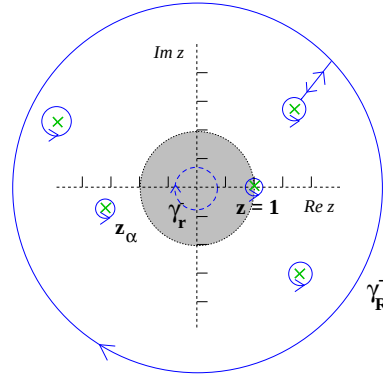
$$\begin{aligned} \Gamma_n &\approx \Gamma_n^{(k)} = \sum_{s_1 s_2 \dots s_n} \tilde{T}_{s_1 s_2 \dots s_k, s_2 \dots s_{k+1}}^{(k)} \tilde{T}_{s_2 \dots s_{k+1}, s_3 \dots s_{k+2}}^{(k)} \dots T_{s_n,} \\ &= \sum_{s_1 s_2 \dots s_k} \sum_{\delta_1 \dots \delta_k} (\tilde{\mathbf{T}}_{(k)}^{n-k})_{s_1 s_2 \dots s_k, \delta_1 \dots \delta_k} \mathcal{M}_{\delta_1 \dots \delta_k}. \end{aligned} \quad (\text{A22.4})$$

Here $\vec{l}$ is the vector of all intervals $\mathcal{M}_i$ at the k th level. It plays the same role as the prefactors a_i in (1.5); in the $n \gg k$ limit, the k th level approximation (A22.4) is dominated by the leading eigenvalue of $\tilde{T}^{(k)}$

$$\Gamma_n^{(k)} \propto [\lambda_{\max}^{(k)}]^n$$

and, as far as the $n \rightarrow \infty$ limit is concerned, the pre-asymptotic intervals $\vec{l}^{(k)}$ contribute only an irrelevant prefactor (unless $\vec{l}^{(k)}$ happens to be normal to the leading eigen-direction of $\tilde{T}^{(k)}$). This method of evaluating sums is familiar from statistical mechanics, whence the designation “transfer operator.”

Figure A22.1: The survival probability Γ_n can be split into contributions from poles (x) and zeros (o) between the small and the large circle and a contribution from the large circle.



A22.1.1 A contour integral formulation



The following observation is sometimes useful, in particular for zeta functions with richer analytic structure than just zeros and poles, as in the case of intermittency (chapter 29): Γ_n , the trace sum (21.22), can be expressed in terms of the dynamical zeta function (22.11)

$$1/\zeta(z) = \prod_p \left(1 - \frac{z^{n_p}}{|\Lambda_p|} \right) . \quad (\text{A22.5})$$

as a contour integral

$$\Gamma_n = \frac{1}{2\pi i} \oint_{\gamma_r^-} z^{-n} \left(\frac{d}{dz} \log \zeta^{-1}(z) \right) dz , \quad (\text{A22.6})$$

where a small contour γ_r^- encircles the origin in negative (clockwise) direction. If the contour is small enough, i.e., it lies inside the unit circle $|z| = 1$, we may write the logarithmic derivative of $\zeta^{-1}(z)$ as a convergent sum over all periodic orbits. Integrals and sums can be interchanged, the integrals can be solved term by term, and the trace formula (21.22) is recovered. For hyperbolic maps, cycle expansions or other techniques provide an analytical continuation of the dynamical zeta function beyond the leading zero; we may therefore deform the original contour into a larger circle with radius R which encircles both poles and zeros of $\zeta^{-1}(z)$, as depicted in figure A22.1. Residue calculus turns this into a sum over the zeros z_α and poles z_β of the dynamical zeta function, that is

$$\Gamma_n = \sum_{|z_\alpha| < R} \frac{1}{z_\alpha^n} - \sum_{|z_\beta| < R} \frac{1}{z_\beta^n} + \frac{1}{2\pi i} \oint_{\gamma_R^-} dz z^{-n} \frac{d}{dz} \log \zeta^{-1} , \quad (\text{A22.7})$$

where the last term gives a contribution from a large circle γ_R^- . It would be a miracle if you still remember this, but in sect. 1.4.3 we interpreted Γ_n as fraction of survivors after n bounces, and defined the escape rate γ as the rate of the find exponential decay of Γ_n . We now see that this exponential decay is dominated by the leading zero or pole of $\zeta^{-1}(z)$.

exercise A22.1

chapter 23

A22.1.2 Dynamical zeta functions for transfer operators



Ruelle's original dynamical zeta function was a generalization of the topological zeta function (18.17) to a function that assigns different weights to different cycles:

chapter 18

$$\zeta(z) = \exp \sum_{n=1}^{\infty} \frac{z^n}{n} \left(\sum_{x_i \in \text{Fix } f^n} \text{tr} \prod_{j=0}^{n-1} g(f^j(x_i)) \right).$$

Here we sum over all periodic points x_i of period n , and $g(x)$ is any (matrix valued) weighting function, where the weight evaluated multiplicatively along the trajectory of x_i .

exercise A22.2

By the chain rule (4.43) the stability of any n -cycle of a 1-dimensional map is given by $\Lambda_p = \prod_{j=1}^n f'(x_i)$, so the 1-dimensional map cycle stability is the simplest example of a multiplicative cycle weight $g(x_i) = 1/|f'(x_i)|$, and indeed - via the Perron-Frobenius evolution operator (19.9) - the historical motivation for Ruelle's more abstract construction.

In particular, for a piecewise-linear map with a finite Markov partition such as the map of example 19.1, the dynamical zeta function is given by a finite polynomial, a straightforward generalization of the topological transition matrix determinant (17.1). As explained in sect. 18.3, for a finite $[N \times N]$ dimensional matrix the determinant is given by

$$\prod_p (1 - t_p) = \sum_{n=1}^N z^n c_n,$$

where c_n is given by the sum over all non-self-intersecting closed paths of length n together with products of all non-intersecting closed paths of total length n .



example A22.1
p. 1012

chapter 18

More generally, piecewise-linear approximations to dynamical systems yield polynomial or rational polynomial cycle expansions, provided that the symbolic dynamics is a subshift of finite type.

We see that the exponential proliferation of cycles so dreaded by quantum chaologists is a bogus anxiety; we are dealing with exponentially many cycles of increasing length and instability, but all that really matters in this example are the stabilities of the two fixed points. Clearly the information carried by the infinity of longer cycles is highly redundant; we have learned in chapter 23 how to exploit this redundancy systematically.

A22.1.3 Perron-Frobenius-Ruelle theory

The transfer operators were introduced by Ruelle [2] as a generalization of the transfer matrices of equilibrium statistical mechanics. For $\beta = 0$ the evolution

operator (20.14) is the kernel of the Perron-Frobenius operator (19.26). For the generalized evolution operators it presumably suffices to refer to Ruelle, who introduced [3] the *Ruelle-Araki operator*

$$(\mathcal{L}^t \circ \rho)(x) = \sum_{y=f^{-t}(x)} e^{A(y)} \rho(y). \quad (\text{A22.8})$$

and who always defines his dynamical zeta functions weighted by multiplicative factors; later papers only rediscover that.

In the mathematical literature, the dominant eigenvalue of the $\mathcal{L}$ operator is related to the averages of the dynamical system by using a theorem of Ruelle (or one of its extensions). Inspired by the statistical mechanics of spin systems, Ruelle extended the Perron-Frobenius theorem for matrices to operators in Banach spaces of bounded functions. He uses the evolution operator introduced in (20.15) with $\beta = 0$ and shows that it can be used to compute the averages of the function h the operator acts on. To do that he assumes that the operator has an isolated largest eigenvalue. With that assumption, he shows that the operator has an adjoint operator acting on the space of measures and that the measure associated with the largest eigenvalue is the invariant measure of the system. A theorem (the Perron-Frobenius-Ruelle theorem) shows that averages of h with this measure can be computed by projecting the function onto the eigenspace of this largest eigenvalue.

The method we use is slightly different. We do not use the action of the operator to compute the averages, but rather rely on the eigenvalue having a few derivatives around $\beta = 0$. The function we want to average is incorporated directly into the definition of the evolution operator; different averages require different operators. The approach is more in the spirit of statistical mechanics, in that we compute a function $s(\beta)$ that is assumed to have a few derivatives with respect to β . That this function exists can be derived from an application of the extensions of Ruelle-Bowen results.

Commentary

Remark A22.1. spectral determinants vs. resolvents. In *30 years in mathematical physics* [1], L. D. Faddeev writes: “For a finite-dimensional operator A the characteristic determinant $\Delta(s) = \det(A - sI)$ satisfies the equation

$$\frac{d}{ds} \ln \Delta(s) = -\text{tr } R(s),$$

where $R(s)$ is the resolvent of the operator A , $R(s) = (A - sI)^{-1}$. This equation can be taken as the basis of the general definition of $\Delta(s)$; however, this requires regularization of the trace of the resolvent, which is less trivial if A has a continuous spectrum. The evolution of the concept of the characteristic determinant is connected with the development of devices for this regularization. From a technical viewpoint I regard $\Delta(s)$ as more convenient than the ζ -function $\zeta(s) = \text{tr } A^{-s}$, popularized in the work of Atiyah and Singer on index theory.”

References

- [1] L. D. Faddeev, “Thirty years in mathematical physics”, *Proc. Steklov Inst. Math.* **176**, 3–28 (1988).
- [2] D. Ruelle, “Statistical mechanics of a one-dimensional lattice gas”, *Commun. Math. Phys.* **9**, 267–288 (1968).
- [3] D. Ruelle, “Zeta-functions for expanding maps and Anosov flows”, *Inv. Math.* **34**, 231–242 (1976).

A22.2 Examples

Example A22.1. A piecewise linear repeller: Due to piecewise linearity, the stability of any n -cycle of the piecewise linear repeller (19.38) factorizes as $\Lambda_{s_1 s_2 \dots s_n} = \Lambda_0^m \Lambda_1^{n-m}$, where m is the total number of times the letter $s_j = 0$ appears in the p symbol sequence, so the traces in the sum (21.24) take the particularly simple form

$$\mathrm{tr} T^n = \Gamma_n = \left(\frac{1}{|\Lambda_0|} + \frac{1}{|\Lambda_1|} \right)^n.$$

The dynamical zeta function (22.10) evaluated by resumming the traces,

$$1/\zeta(z) = 1 - z/|\Lambda_0| - z/|\Lambda_1|, \quad (\text{A22.9})$$

is indeed the determinant $\det(1 - zT)$ of the transfer operator (19.40), which is almost as simple as the topological zeta function (18.24).

[exercise A22.3](#)

[return to p. 1010](#)

A22.3 Exercises

A22.1. **Contour integral for survival probability.** Perform explicitly the contour integral appearing in (A22.6).

A22.2. **General weights.** (easy) Let f^t be a flow and $\mathcal{L}^t$ the operator

$$\mathcal{L}^t g(x) = \int dy \delta(x - f^t(y)) w(t, y) g(y)$$

where w is a weight function. In this problem we will try and determine some of the properties w must satisfy.

(a) Compute $\mathcal{L}^s \mathcal{L}^t g(x)$ to show that

$$w(s, f^t(x)) w(t, x) = w(t + s, x).$$

(b) Restrict t and s to be integers and show that the most general form of w is

$$w(n, x) = g(x) g(f(x)) g(f^2(x)) \cdots g(f^{n-1}(x)),$$

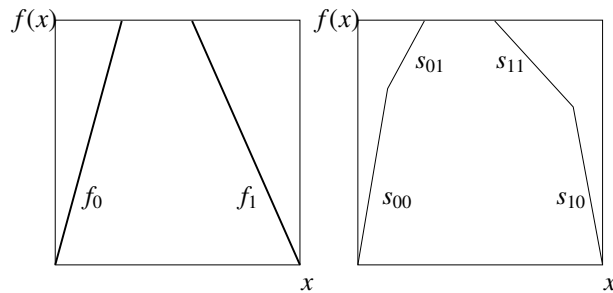
for some g that can be multiplied. Could g be a function from $\mathbb{R}^{n_1} \mapsto \mathbb{R}^{n_2}$? ($n_i \in \mathbb{N}$.)

A22.3. **Dynamical zeta functions.** (easy)

(a) Evaluate in closed form the dynamical zeta function

$$1/\zeta(z) = \prod_p \left(1 - \frac{z^{n_p}}{|\Lambda_p|} \right)$$

for the piecewise-linear map (19.38) with the left branch slope Λ_0 , the right branch slope Λ_1 .



(b) What if there are four different slopes s_{00} , s_{01} , s_{10} , and s_{11} instead of just two, with the preimages of the gap adjusted so that junctions of branches s_{00} , s_{01} and s_{11} , s_{10} map in the gap in one iteration? What would the dynamical zeta function be?

Appendix A24

Deterministic diffusion

BASIC NOTIONS of discretizing continuum are introduced: discretized fields on lattices, lattice derivatives, lattice Laplacians. Invariance of a given theory under (discrete) translations motivates us to consider periodic lattices, and use the eigenmodes of translation generators to diagonalise (discrete Fourier transformations) non-local operators, such as Laplacians, and invert them.

We then use these tools to study in sect. A24.6 some of the simplest examples of deterministic systems that exhibit “deterministic diffusion,” the sawtooth and cat maps.

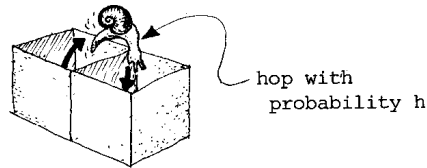
A24.1 Wanderings of a drunken snail

Statistical mechanics is formulated in a Euclidean world in which there is no time, just space. Q. What do we mean by ‘propagation’ in such a space?

We have no idea what the structure of our space on distances much shorter than interatomic might be. The very space-time might be discrete rather than continuous, or it might have geometry different from the one we observe at the accessible distance scales. The formalism we use should reflect this ignorance. We deal with this problem by coarse-graining the space into small hypercubes (‘cells’), and requiring that our theory be insensitive to distances comparable to or smaller than the cell size.

Our next problem is that we have no idea why there are “particles,” and why or how they propagate. The most we can say is that there is some probability that a particle hops from one cell to the next cell. This stepping probability is a phenomenological, coarse-graining parameter which - in the continuum limit - is related to the “mass” of the particle.

Assume that initially the particle is confined to a cell at location $z = (z_1, z_2, \dots, z_d)$, and that the particle is a ‘scalar’, i.e., it has no further internal degrees of freedom, such as spin or color. What is it like to be free? A free particle exists only in itself and for itself; it neither sees nor feels the others; it is, in this chilly sense, free. But if it is not at once paralyzed by the vast possibilities opened to it, it soon becomes perplexed by the problems of realizing any of them alone. Born free, it is constrained by the very lack of constraint. Sitting in its cell, it is faced by a choice of doing nothing (s = stopping probability) or stepping into any of the $2d$ neighboring cells (h = stepping probability):



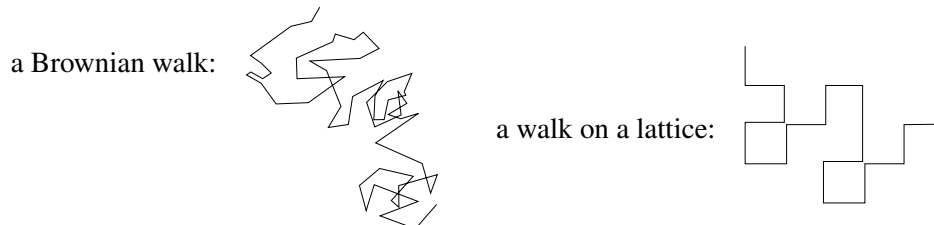
The number of neighboring cells defines the dimension of the space. The stepping and stopping probabilities are related by the probability conservation:

$$1 = s + 2dh, \quad d = \text{the dimension of spacetime} . \quad (\text{A24.1})$$

Taking the stepping probability h to be the same in all directions means that we have assumed that the space is *isotropic*. Taking the stopping probability s to be the same everywhere means that we have assumed that the space is *homogeneous*. Otherwise the propagation is not free, but is constrained by some external geometry. This can either mean that the space is infinite, or that it is compact and periodic (a torus; a Lie group manifold).

The isotropy and homogeneity assumptions imply that our theory should be invariant under lattice rotations and discrete translations in units of the lattice spacing. If the theory we are discretizing is insensitive to the very short and very long distances, it is said to be well behaved in the ‘ultraviolet’ and ‘infrared’ limits.

While counting continuum Brownian paths might be tricky, counting them in a discretized space is a breeze. Divide the space into little d -dimensional hypercubes (‘cells’, ‘tiles’). Let the snail start in the cell z and hop along until it stops in the cell z' .



The probability of this path is $h^\ell s$, where ℓ is the number of steps. The total probability that a particle wanders from the z th cell and stops in the z' th cell is the

sum of probabilities associated with all distinct paths between them:

$$\Delta_{z'z} = s \sum_{\ell=0}^{\infty} h^{\ell} N_{z'z}(\ell), \quad (\text{A24.2})$$

where $N_{z'z}(\ell)$ is the number of all paths of ℓ steps connecting the lattice site z to site z' . In order to compute $N_{z'z}(\ell)$, define a shift operator and its inverse (also its transpose)

$$(r_j)_{z'z} = \delta_{z'+\mathbf{e}_j, z}, \quad (r_j^{-1})_{z'z} = \delta_{z', z+\mathbf{e}_j}, \quad (\text{A24.3})$$

where $\mathbf{e}_j$ is the unit step in direction j . If a particle is introduced into the cell z , with probability $\rho_z^0 = \delta_{z'z}$, the shift operator moves it into a neighboring cell:

$$(r_j \rho^0)_{z'} = \delta_{z'+\mathbf{e}_j, z} \quad \Rightarrow \quad \begin{array}{c} z' \\ \swarrow \searrow \\ z' + \mathbf{e}_j \end{array} \quad (\text{A24.4})$$

The lattice diffusion operator

$$\mathcal{D}_{z'z} = \sum_{j=1}^d (r_j + r_j^{-1})_{z'z} \quad (\text{A24.5})$$

generates all steps of length 1:

$$\mathcal{D}\rho^0 = \begin{array}{ccc} & \circ & \\ & \swarrow \searrow & \\ 1 \circ & & \circ 1 \\ & \swarrow \searrow & \\ & \circ & \\ & \swarrow \searrow & \\ 1 \circ & & \circ 1 \end{array}, \quad (\text{A24.6})$$

(the examples are drawn in two dimensions; the lattice does not look square as this is meant to be a sidewise view from above the lattice). The paths of length 2 are counted by

$$\mathcal{D}^2 \rho^0 = \begin{array}{ccccc} & & \circ & & \\ & & \swarrow \searrow & & \\ & \circ & & \circ & \\ & \swarrow \searrow & & \swarrow \searrow & \\ 1 \circ & & \circ & & \circ 1 \\ & \swarrow \searrow & & \swarrow \searrow & \\ & \circ & & \circ & \\ & \swarrow \searrow & & \swarrow \searrow & \\ 2 \circ & & \circ & & \circ 2 \\ & \swarrow \searrow & & \swarrow \searrow & \\ & \circ & & \circ & \\ & \swarrow \searrow & & \swarrow \searrow & \\ 1 \circ & & \circ & & \circ 1 \end{array}. \quad (\text{A24.7})$$

There are 4 ways of returning, 2 ways of reaching a diagonal corner, and so on. Note –and this is the key observation– that the z' th component of the vector $\mathcal{D}^{\ell} \rho^0$ counts the number of paths of length ℓ connecting the final z' cell to the initial z th cell. The probability that the particle stops in the z' th cell is given by the ‘path sum’

$$\rho_{z'} = s \sum_{\ell=0}^{\infty} h^{\ell} (\mathcal{D}^{\ell} \rho^0)_{z'}, \quad (\text{A24.8})$$

the sum of all paths connecting the two cells. $\rho_{z'}$ is the probability that the particle introduced into the cell z with probability ρ_z^0 ends in the cell z' . According to (A24.6), (A24.7), the numbers of paths originating in cell z are

$$\sum_z (\mathcal{D} \rho^0)_z = 2d, \quad \sum_z (\mathcal{D}^2 \rho^0)_z = (2d)^2, \quad \dots$$

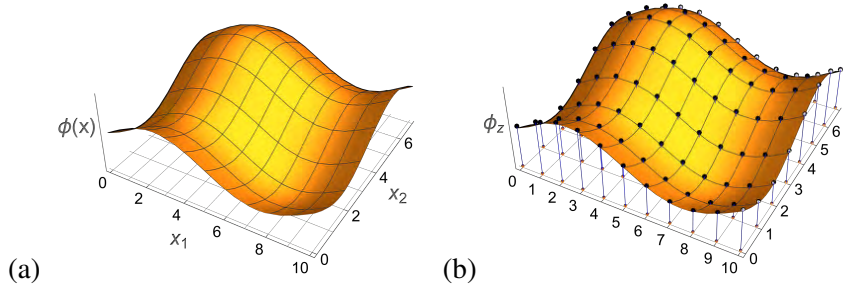


Figure A24.1: Discretization of a field over two-dimensional spacetime. (a) A periodic scalar field configuration $\phi(x)$ over a primitive cell of spatial period L , temporal period T , plotted as a function of continuous coordinates $x \in \mathbb{R}^2$. (b) The corresponding discretized field configuration (A24.13) over $[10 \times 7]_0$ primitive cell, with the field value ϕ_z at the lattice site $z \in \mathbb{Z}^2$ indicated by a dot. (Han Liang [10].)

so, by the stepwise probability conservation (A24.1), ρ_z is a correctly normalized probability field (probability density on lattice site z),

$$\sum_z \rho_z = s(1 + 2dh + (2d)^2 h^2 + \dots) = \frac{s}{1 - 2dh} = 1. \quad (\text{A24.9})$$

Actually, the powers of the hopping operator (A24.5) also form a geometric series, which we can sum up like we would sum up a matrix series,

$$\rho_z = \Delta \rho^0, \quad \Delta_{z'z} = \left(\frac{s}{1 - h\mathcal{D}} \right)_{z'z}, \quad (\text{A24.10})$$

yielding an explicit formula for the Euclidean free scalar particle propagator (A24.2).

A. The probability $\Delta_{z'z}$ that a drunken snail starting out in the cell z ends up in the cell z' is given by the sum of all paths starting in z and ending in z' .

A24.1.1 Lattice field theory

Let us set up some ‘lattice field theory’ notation, conventional in solid state physics and quantum field theory. We’ll return to the evaluation of the Euclidean free massive scalar particle propagator (A24.10) in sect. A24.4.5.

In a d -dimensional hypercubic discretization of a Euclidean space, the d continuous Euclidean coordinates $x \in \mathbb{R}^d$ are replaced by a hypercubic integer lattice

$$\mathcal{L} = \left\{ \sum_{j=1}^d z_j \mathbf{e}_j \mid z \in \mathbb{Z}^d \right\}, \quad \mathbf{e}_j \in \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_d\}, \quad (\text{A24.11})$$

spanned by a set of orthogonal unit vectors $\mathbf{e}_j$, with lattice spacing $a_j = |\mathbf{e}_j| = \Delta x_j$ along the direction of unit vector $\mathbf{e}_j$. We shall use lattice units, always setting $a_j = 1$. A field $\phi(x)$ over d continuous coordinates x_j is represented by a discrete array of field values over lattice sites

$$\phi_z = \phi(x), \quad x_j = a_j z_j = \text{lattice site}, \quad z \in \mathbb{Z}^d, \quad (\text{A24.12})$$

as sketched in figure A24.1. A lattice *field configuration* is a d -dimensional array of field values (in what follows, illustrative examples will be presented in one or two spatiotemporal dimensions)

$$\Phi = \begin{array}{ccccccc} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \phi_{-2,1} & \phi_{-1,1} & \phi_{0,1} & \phi_{1,1} & \phi_{2,1} & \cdots \\ \cdots & \phi_{-2,0} & \phi_{-1,0} & \phi_{0,0} & \phi_{1,0} & \phi_{2,0} & \cdots \\ \cdots & \phi_{-2,-1} & \phi_{-1,-1} & \phi_{0,-1} & \phi_{1,-1} & \phi_{2,-1} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{array}. \quad (\text{A24.13})$$

A field configuration is a *point* in system's *state space*

$$\mathcal{M} = \{ \Phi \mid \phi_z \in \mathbb{R}, \quad z \in \mathbb{Z}^d \}, \quad (\text{A24.14})$$

the totality of states Φ , given by all possible values of site fields, where ϕ_z can be a single scalar field, or a multiplet of real or complex fields.

While we refer here to such discretization as a lattice field theory, the lattice might arise naturally from a many-body setting with the nearest neighbors interactions, such as many-body quantum chaos models, with a multiplet of fields at every site.

A24.1.2 Periodic field configurations

Q. What are these ‘operators’, such as the shift operator (A24.3)? The hypercubic lattice $z \in \mathbb{Z}^d$ is infinite in all directions, so these are not matrices, they are infinite-dimensional objects. Our bridge to this infinity will be to tile the infinite lattice by repeats of finite tiles.

To get a grip on these ‘operators’, let’s start small, and make the lattice 1-dimensional and finite, a chain of L lattice sites. For a hyper-cubic lattice the translations in different directions commute, $r_j r_i = r_i r_j$, so it is sufficient to understand the action of a shift on a 1-dimensional lattice. So we can write down r (A24.3) in its full $[L \times L]$ matrix glory. Writing the finite lattice shift operator as an ‘upper shift’ matrix

$$r = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ 0 & & & & 0 \end{pmatrix},$$

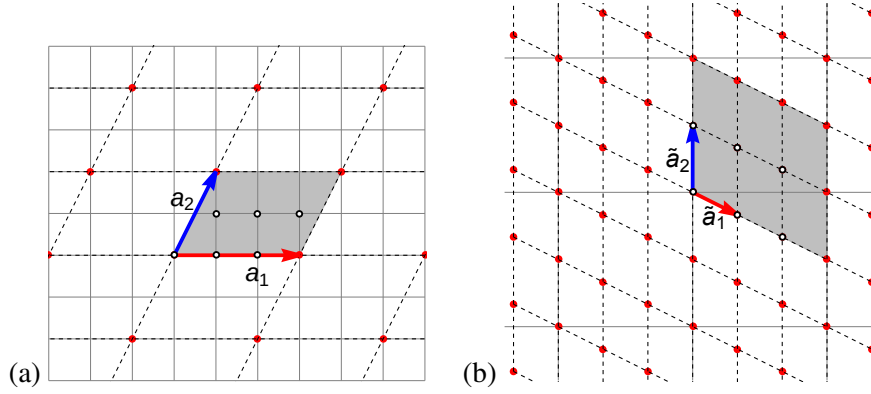


Figure A24.2: (a) The intersection points z of the light grey lines form the integer square lattice (A24.12). The primitive vectors $\mathbf{a}_1 = (3, 0)$ and $\mathbf{a}_2 = (1, 2)$ form the primitive cell $\mathbb{A} = [3 \times 2]_1$ (see (A24.17)), whose translations tile the Bravais lattice $\mathcal{L}_{\mathbb{A}}$ (red points). (b) The intersection points k of the light grey lines form the reciprocal square lattice. Translations of reciprocal primitive vectors $\tilde{\mathbf{a}}_1$ and $\tilde{\mathbf{a}}_2$ generate the reciprocal lattice $\mathcal{L}_{\tilde{\mathbb{A}}}$ (red points). (Shaded) The reciprocal primitive cell $\tilde{\mathbb{A}}$. A wave vector outside this region is equivalent to a wave vector within it by a reciprocal lattice translation. Note that the number of lattice sites within the reciprocal primitive cell $\tilde{\mathbb{A}}$ equals the number of sites within the spatiotemporal primitive cell $\mathbb{A}$. (Han Liang [10].)

is no good, as r so defined is nilpotent, and after L steps the particle marches off the lattice edge, and nothing is left, $r^L = 0$. The smart way to approximate an infinite lattice by a finite one is to insist that the discretization preserve the translational invariance, and represent the *periodic* lattice operator. On a *periodic lattice* every point is equally far from the ‘boundary’ $L/2$ steps away, the ‘surface’ effects are equal for all points, and the shift operator $r_{z'z} = \delta_{z'+1,z}$ acts as a cyclic permutation matrix

$$r = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ 1 & & & & 0 \end{pmatrix}, \quad (\text{A24.15})$$

with ‘1’ in the lower left corner assuring periodicity. Now, let's go back to the infinite lattice, by tiling the infinite lattice by repeats of a finite periodic field configuration.

A lattice field configuration is $\mathcal{L}_{\mathbb{A}}$ -periodic if

$$\phi_{z+r} = \phi_z \quad (\text{A24.16})$$

for any discrete translation $r = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + \cdots n_d \mathbf{a}_d$ in the *Bravais lattice*

$$\mathcal{L}_{\mathbb{A}} = \left\{ \sum_{j=1}^d n_j \mathbf{a}_j \mid n_j \in \mathbb{Z} \right\}, \quad (\text{A24.17})$$

where the $[d \times d]$ matrix $\mathbb{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_d]$ formed from primitive lattice vectors $\{\mathbf{a}_j\}$ defines a d -dimensional *primitive cell* [3, 8] (see figure A24.2 (a)).

Primitive cell $\mathbb{A}$ field configuration lattice-site fields (A24.13) take values in the $V_{\mathbb{A}}$ -dimensional state space

$$\mathcal{M}_{\mathbb{A}} = \{\Phi \mid \phi_z \in \mathbb{R}, z \in \mathbb{A}\}. \quad (\text{A24.18})$$

The *volume* of Bravais lattice $\mathcal{L}_{\mathbb{A}}$ equals the number of lattice sites $z \in \mathbb{A}$ within the primitive cell (see figure A24.2 (a)):

$$V_{\mathbb{A}} = |\text{Det } \mathbb{A}|. \quad (\text{A24.19})$$

For example, repeats of the $V_{\mathbb{A}} = 15$ -dimensional $[5 \times 3]$ primitive cell field configuration

$$\Phi = \begin{bmatrix} \phi_{-2,1} & \phi_{-1,1} & \phi_{0,1} & \phi_{1,1} & \phi_{2,1} \\ \phi_{-2,0} & \phi_{-1,0} & \phi_{0,0} & \phi_{1,0} & \phi_{2,0} \\ \phi_{-2,-1} & \phi_{-1,-1} & \phi_{0,-1} & \phi_{1,-1} & \phi_{2,-1} \end{bmatrix} \quad (\text{A24.20})$$

tile periodically the doubly-infinite state space (A24.13).

A. While an operator, such as the shift operator (A24.3), is an infinite-dimensional object, periodic tilings of spacetime enable us to treat operators as though they were finite-dimensional matrices.

A24.1.3 Orbits

These periodic field configurations are described by a finite number $V_{\mathbb{A}}$ of field values.

Q. Does the way they are laid out on the lattice matter?

Consider a one-dimensional primitive cell $\mathbb{A}$, defined by a single primitive vector $\mathbf{a}_1 = n$ in (A24.17). One-lattice-spacing shift operator (A24.15) is a cyclic permutation operator that translates a field configuration by one lattice site,

$$\begin{aligned} \Phi &= [\phi_0 \phi_1 \phi_2 \phi_3 \cdots \phi_{n-1}] \\ r\Phi &= [\phi_1 \phi_2 \phi_3 \cdots \phi_{n-1} \phi_0], \quad \text{or} \quad (r\Phi)_z = \phi_{z+1}, \\ &\dots \\ r^{n-1}\Phi &= [\phi_{n-1} \phi_0 \phi_1 \phi_2 \cdots \phi_3], \\ r^n\Phi &= [\phi_0 \phi_1 \phi_2 \phi_3 \cdots \phi_{n-1}], \quad \text{so} \quad r^n\Phi = \Phi. \end{aligned} \quad (\text{A24.21})$$

While each field configuration $r^j\Phi$ might be a distinct point in the primitive cell's state space (A24.18), they are equivalent, in the sense that all consist of the same set of lattice site fields $\{\phi_z\}$, up to a cyclic relabelling of lattice sites.

In this way actions of a group of relabelling permutations $g \in G$ on field configurations over a multi-periodic primitive cell $\mathbb{A}$ foliate the state space into a union

$$\mathcal{M}_{\mathbb{A}} = \{\Phi\} = \cup_p \mathcal{M}_p \quad (\text{A24.22})$$

of *orbits*,

$$\mathcal{M}_p = \{g \Phi_p \mid g \in G\} \quad (\text{A24.23})$$

each a set of equivalent field configurations, labelled p , or perhaps by Φ_p , one of the configurations in the set. Each orbit is a fixed point of G , as any element $g\mathcal{M}_p = \mathcal{M}_p$ simply permutes the configurations within the set. The number of distinct field configurations in the orbit is known as the *index* of orbit $\mathcal{M}_p$. It can be as large as $|G|$, the number of elements in G , or as small as 1, if the field configuration is $\phi_z = \phi$, a constant (or a ‘steady state’).

A. While an operator, such as the shift operator (A24.3), is an infinite-dimensional object, periodic tilings of spacetime enable us to treat operators as finite-dimensional matrices.

A24.2 Lattice derivatives

In order to measure spatial variations of fields, we need to define *lattice derivatives*.

Consider a function $\phi(x)$ discretized on a d -dimensional lattice (A24.12). Each set of field values ϕ -a vector ϕ_z - is a possible *field configuration*.

Assume the lattice is a hypercubic integer lattice (A24.11), and let $\{\mathbf{e}_j\}$ be the unit lattice cell vectors pointing along the d positive directions. The forward *lattice derivative* is then

$$(\partial_j \phi)_z = \frac{\phi(x + a\mathbf{e}_j) - \phi(x)}{a} = \frac{\phi_{z+\mathbf{e}_j} - \phi_z}{a}. \quad (\text{A24.24})$$

The backward *lattice derivative* is its transpose

$$(\partial_j \phi)_z^\top = \frac{\phi(x - a\mathbf{e}_j) - \phi(x)}{a} = \frac{\phi_{z-\mathbf{e}_j} - \phi_z}{a}. \quad (\text{A24.25})$$

Anything else with the correct $a \rightarrow 0$ limit would do, but this is the simplest choice. We can rewrite the lattice derivative as a linear operator, by introducing the *shift operator* (or *stepping operator*) (A24.3) in the direction j . The partial lattice derivative (A24.24) can now be written as a multiplication by the operator:

$$\partial_j \phi_z = \frac{1}{a} (r_j - \mathbf{1})_{zz'} \phi_{z'}.$$

In the 1-dimensional case the $[L \times L]$ matrix representation of the lattice derivative is:

$$\partial = \frac{1}{a} \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & -1 & 1 & \\ & & & \ddots & \\ & 1 & & & -1 \end{pmatrix}. \quad (\text{A24.26})$$

To belabor the obvious: On a finite lattice of L points a derivative is simply a finite $[L \times L]$ matrix. Continuum field theory is a world in which the lattice is so fine that it looks smooth to us. Whenever someone calls something an “operator,” think “matrix.” For finite-dimensional spaces a linear operator *is* a matrix; things get subtler for infinite-dimensional spaces.

Looking forward, the continuum limit of this operator will be related to what field theorists call particle ‘momentum’, so we find it convenient to refer to the forward lattice difference operator in j th lattice direction as the ‘lattice momentum’ operator

$$p_j = r_j - \mathbb{1}, \quad (\text{A24.27})$$

where the lattice spacing is set to $a = 1$ (easily reinstated, if needed).

A24.2.1 Lattice Laplacian

I’m the field ϕ_z on lattice site z .
Q. What’s up with my neighbors?

In lattice field theory ‘locality’ means that a field at site z interacts only with its neighbors. To keep the exposition as simple as possible, we shall -for time being- treat the spatial and temporal directions on equal footing, with the graph Laplace operator [7, 12, 18, 19]

$$\square \phi_z = \sum_{z' : \|z' - z\|=1} (\phi_{z'} - \phi_z) \quad \text{for all } z, z' \in \mathbb{Z}^d \quad (\text{A24.28})$$

comparing the field on lattice site z to its $2d$ nearest neighbors.

On the d -dimensional hyper-cubic lattice, Laplacian is the centered, reflection symmetric second lattice derivative,

$$\square = \sum_{j=1}^d (r_j - 2\mathbb{1} + r_j^{-1}). \quad (\text{A24.29})$$

For example, the two-dimensional square lattice Laplace operator is given by

$$\square = r_1 + r_2 - 4\mathbb{1} + r_2^{-1} + r_1^{-1}, \quad (\text{A24.30})$$

and in the 1-dimensional finite periodic chain case the $[L \times L]$ matrix representation of the lattice Laplacian is

$$\square = \frac{1}{a^2} \begin{pmatrix} -2 & 1 & & & 1 \\ 1 & -2 & 1 & & \\ & 1 & -2 & 1 & \\ & & 1 & \ddots & \\ 1 & & & 1 & -2 \end{pmatrix}. \quad (\text{A24.31})$$

The lattice Laplacian measures the second variation of a field ϕ_z across three neighboring sites: it is spatially *non-local*. You can easily check that it does what the second derivative is supposed to do by applying it to a parabola restricted to the lattice, $\phi_z = \phi(az)$, where $\phi(az)$ is defined by the value of the continuum function $\phi(x) = \phi^2$ at the lattice point $x = az$.

We can write it in terms of the ‘lattice momentum’ operator (A24.27), and relate it to $\mathcal{D}$, the hopping operator (A24.5),

$$\begin{aligned} \square &= - \sum_{j=1}^d \mathbf{p}_j^\top \mathbf{p}_j = - \sum_{j=1}^d (r_j^{-1} - \mathbf{1})(r_j - \mathbf{1}) \\ &= \frac{1}{a^2} (\mathcal{D} - 2d\mathbf{1}) . \end{aligned} \quad (\text{A24.32})$$

The Euclidean free scalar particle propagator (A24.10) can thus be written as

$$\Delta = \frac{1}{\mathbf{1} - \frac{\hbar}{s} a^2 \square} . \quad (\text{A24.33})$$

In what follows it will be convenient to reinterpret and rescale this drunken-walk propagator Δ , and consider instead the “free field action” of form

$$S[\phi] = -\frac{1}{2} \phi^\top \cdot M^{-1} \cdot \phi . \quad (\text{A24.34})$$

where the “free” or “bare” massive scalar propagator M is parametrized as

$$M = \frac{1}{p^2 + \mu^2} . \quad (\text{A24.35})$$

What this parametrization says is that the mass squared m^2 of the Euclidean scalar particle is proportional to $m^2 \sim s/\hbar$: the heavier the particle, the less likely it is to hop, the more likely is it to stop.

A. Laplacian measures local curvature, how much I stick out among my neighbors.

A24.2.2 Integration by parts



(Please skip this and all other bits marked ‘cyclist’ on the first reading - not essential at this time.)

In the continuum, integration by parts moves ∂ around,

$$\int [dx] \phi^\top \partial^2 \phi \rightarrow - \int [dx] \partial \phi^\top \cdot \partial \phi ;$$

on a lattice this amounts to a matrix transposition

$$\left[(r_j - \mathbf{1})\phi\right]^\top \cdot \left[(r_j - \mathbf{1})\phi\right] = \phi^\top \cdot (r_j^{-1} - \mathbf{1})(r_j - \mathbf{1})\phi.$$

If you are wondering where the “integration by parts” minus sign is, it is there in discrete case as well. It comes from the identity

$$\partial^\top = \frac{1}{a}(r^{-1} - \mathbf{1}) = -r^{-1}\frac{1}{a}(r - \mathbf{1}) = -r^{-1}\partial.$$

Integrating by parts is now “summing by parts.” Let a_i and b_i be n -periodic vectors, and $(\partial a)_i = a_i - a_{i-1}$ be the difference operator. Then

$$\sum_{i=1}^n (\partial a)_i b_i = - \sum_{i=1}^n a_i (\partial b)_{i+1}. \quad (\text{A24.36})$$

A24.2.3 Inverting the Laplacian

Evaluation of perturbative corrections requires that we come to grips with the “free” or “bare” propagator M . While the Laplacian is a simple difference operator (A24.31), the propagator is a messier object. A way to compute is to start expanding the propagator M as a power series in the Laplacian

$$M = \frac{1}{p^2 + \mu^2} = \frac{1}{m^2} \sum_{k=0}^{\infty} \frac{1}{m^{2k}} \square^k. \quad (\text{A24.37})$$

As $\square$ is a finite matrix, the expansion is convergent for sufficiently large m^2 . To get a feeling for what is involved in evaluating such series, evaluate $\square^2$ in the 1-dimensional case:

$$\square^2 = \frac{1}{a^4} \begin{pmatrix} 6 & -4 & 1 & & & 1 & -4 \\ -4 & 6 & -4 & 1 & & & 1 \\ 1 & -4 & 6 & -4 & 1 & & \\ & 1 & -4 & \ddots & & & 1 \\ 1 & & & & & 6 & -4 \\ -4 & 1 & & & 1 & -4 & 6 \end{pmatrix}. \quad (\text{A24.38})$$

What $\square^3$, $\square^4$, $\dots$ contributions look like is now clear; as we include higher and higher powers of the Laplacian, the propagator matrix fills up; while the *inverse* propagator is differential operator connecting only the nearest neighbors, the propagator is integral, *non-local* operator, connecting every lattice site to any other lattice site. Due to the periodicity, these are all Toeplitz matrices, meaning that each successive row is a one-step cyclic shift of the preceding one. In statistical mechanics, M is the (bare) 2-point correlation. In quantum field theory, it is called a propagator.

These matrices can be evaluated as is, on the lattice, and sometime it is evaluated this way, but in case at hand a wonderful simplification follows from the observation that the lattice action is translationally invariant. We show how this works in sect. A24.3.

A24.3 Periodic lattices

Our task now is to transform M into a form suitable to explicit evaluation.

Consider the effect of a lattice translation $\phi \rightarrow r\phi$ on the matrix polynomial

$$S[r\phi] = -\frac{1}{2}\phi^\top (r^\top M^{-1} r) \phi.$$

As M^{-1} is constructed from r and its inverse, M^{-1} and r commute, and $S[\phi]$ is invariant under translations,

$$S[r\phi] = S[\phi] = -\frac{1}{2}\phi^\top \cdot \frac{1}{M} \cdot \phi. \quad (\text{A24.39})$$

If a function defined on a vector space commutes with a linear operator r , then the eigenvalues of r can be used to decompose the ϕ vector space into invariant subspaces. For a hyper-cubic lattice the translations in different directions commute, $r_j r_i = r_i r_j$, so it is sufficient to understand the spectrum of the 1-dimensional shift operator (A24.15).

To develop a feeling for how this reduction to invariant subspaces works in practice, let us proceed cautiously, by expanding the scope of our deliberations to a lattice consisting of 2 points.



example A24.1
p. 1039

A24.4 Discrete Fourier transforms

Let us generalize this reduction to a 1-dimensional periodic lattice with L sites.

A24.4.1 Reciprocal primitive cell

Translation invariance of orbits suggests reformulating the theory in a discrete Fourier basis, a discretization approach that goes all the way back to Hill's 1886 paper [14].

The n consecutive shifts (A24.21) return a period- n field configuration to itself, so acting on an one-dimensional periodic primitive cell, shift operator satisfies the characteristic equation

$$r^n - \mathbb{1} = \prod_{m=0}^{n-1} (r - e^{ik} \mathbb{1}) = 0, \quad (\text{A24.40})$$

with the n -th roots of unity eigenvalues $\{e^{ik}\}$ indexed by integers m

$$k = \frac{2\pi}{n} m, \quad m = 0, 1, \dots, n-1, \quad (\text{A24.41})$$

and n eigenvectors

$$r\varphi(k) = e^{ik}\varphi(k), \quad [\varphi(k)]_z = e^{ikz}. \quad (\text{A24.42})$$

The shift (A24.21)

$$[r\varphi(k)]_z = [\varphi(k)]_{z+1} = e^{ik(z+1)} = e^{ik}[\varphi(k)]_z$$

acts by rotating the eigenvector's overall phase.

Wave numbers k form a one-dimensional *reciprocal lattice*,

$$\mathcal{L}_{\tilde{\mathbf{A}}} = \{m\tilde{\mathbf{a}}_1 \mid m \in \mathbb{Z}\}, \quad \tilde{\mathbf{a}}_1 \cdot \mathbf{a}_1 = 2\pi,$$

with the primitive reciprocal lattice vector $\tilde{\mathbf{a}}_1 = 2\pi/n$, and the reciprocal primitive cell –the interval $[0, 2\pi)$ – that contains n distinct wave numbers (A24.41).

Each application of r translates the lattice one step; in L steps the lattice is back in the original state

$$r^L = \mathbf{1} \quad \begin{array}{c} \circ^5 \quad \circ^4 \quad \circ^3 \\ \circ^2 \\ \circ^1 \\ \circ^0 \\ \circ^{N-1} \\ \circ^{N-2} \\ \circ^0 \\ \circ^1 \\ \circ^2 \\ \circ^3 \\ \circ^4 \\ \circ^5 \end{array}, \quad (\text{A24.43})$$

so the eigenvalues of r are the L distinct L th roots of unity

$$r^L - \mathbf{1} = \prod_{k=0}^{L-1} (r - \omega^k \mathbf{1}) = 0, \quad \omega = e^{i2\pi/L}. \quad (\text{A24.44})$$

As the eigenvalues are all distinct and L in number, the space is decomposed into L 1-dimensional subspaces.

A24.4.2 Projection operators



(Please skip this and all other bits marked ‘cyclist’ on the first reading - not essential at this time.)

The general theory (expounded in appendix A10.2) associates with the k th eigenvalue of r a projection operator that projects a state ϕ onto k th eigenvector of r ,

$$P_k = \prod_{j \neq k} \frac{r - \omega^j \mathbf{1}}{\omega^k - \omega^j}. \quad (\text{A24.45})$$

A factor $(r - \omega^j \mathbf{1})$ kills the j th eigenvector φ_j component of an arbitrary vector in expansion $\phi = \dots + \tilde{\phi}_j \varphi_j + \dots$. The above product kills everything but the eigen-direction φ_k , and the factor $\prod_{j \neq k} (\omega^k - \omega^j)$ ensures that P_k is normalized as a projection operator. The set of the projection operators is complete,

$$\sum_k P_k = \mathbf{1}, \quad (\text{A24.46})$$

and orthonormal

$$P_k P_j = \delta_{kj} P_k \quad (\text{no sum on } k). \quad (\text{A24.47})$$

In the case of discrete translational invariance, or cyclic group C_L , it is customary to write out the projection operator (A24.45) as a character-weighted sum, see example A24.2.

As any matrix function $M = M(r)$ of the translation generator r takes a scalar value on the k th subspace,

$$M(r) P_k = M(\omega^k) P_k, \quad (\text{A24.48})$$

the projection operators *diagonalize* the matrix M , $P_j M(r) P_k = M(\omega^k) P_k \delta_{jk}$.

The $[L \times L]$ projection operator matrix elements can be expressed in terms of the eigenvectors (A24.51), (A24.52) as

$$(P_k)_{\ell\ell'} = (\varphi_k)_\ell (\varphi_k^\dagger)_{\ell'} = \frac{1}{L} e^{i\frac{2\pi}{L}(\ell-\ell')k}, \quad (\text{no sum on } k). \quad (\text{A24.49})$$

The completeness (A24.46) follows from (A24.54), and the orthonormality (A24.47) from (A24.53).

$\tilde{\phi}_k$, the projection of the L -dimensional state (i.e., vector) ϕ on the k th subspace is given by

$$\begin{aligned} (P_k \cdot \phi)_\ell &= \tilde{\phi}_k (\varphi_k)_\ell, \quad (\text{no sum on } k) \\ \tilde{\phi}_k &= \varphi_k^\dagger \cdot \phi = \frac{1}{\sqrt{L}} \sum_{\ell=0}^{L-1} e^{-i\frac{2\pi}{L}k\ell} \phi_\ell \end{aligned} \quad (\text{A24.50})$$

A24.4.3 Eigenvectors of the translation operator

While constructing explicit eigenvectors is usually not a the best way to fritter one's youth away, as choice of basis is largely arbitrary, and all of the content of the theory is in the projection operators (see appendix A10.2), in case at hand the eigenvectors are so simple that we can construct and verify the solutions of the eigenvalue condition

$$r \varphi_k = \omega^k \varphi_k \quad (\text{A24.51})$$

by hand:

$$\frac{1}{\sqrt{L}} \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & 0 & 1 & \\ & & & \ddots & \\ & & & & 0 & 1 \\ 1 & & & & & 0 \end{bmatrix} \begin{bmatrix} 1 \\ \omega^k \\ \omega^{2k} \\ \omega^{3k} \\ \vdots \\ \omega^{(L-1)k} \end{bmatrix} = \omega^k \frac{1}{\sqrt{L}} \begin{bmatrix} 1 \\ \omega^k \\ \omega^{2k} \\ \omega^{3k} \\ \vdots \\ \omega^{(L-1)k} \end{bmatrix}$$

In words: the cyclic translation generator r shifts all components by one, and the original vector is recovered by factoring out the common factor ω^k . The $1/\sqrt{L}$ factor normalizes φ_k to a complex unit vector,

$$\begin{aligned}\varphi_k^\dagger \cdot \varphi_k &= \frac{1}{L} \sum_{k=0}^{L-1} 1 = 1, \quad (\text{no sum on } k) \\ \varphi_k^\dagger &= \frac{1}{\sqrt{L}} \left(1, \omega^{-k}, \omega^{-2k}, \dots, \omega^{-(L-1)k} \right).\end{aligned}\quad (\text{A24.52})$$

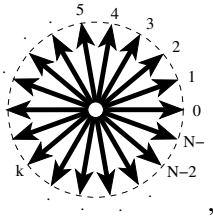
The eigenvectors are orthonormal

$$\varphi_k^\dagger \cdot \varphi_j = \delta_{kj}, \quad (\text{A24.53})$$

as the explicit evaluation of $\varphi_k^\dagger \cdot \varphi_j$ yields the *Kronecker (circular) delta function for a periodic lattice*

$$\delta_{kj} = \frac{1}{L} \sum_{\ell=0}^{L-1} e^{i \frac{2\pi}{L} (k-j)\ell}. \quad (\text{A24.54})$$

The sum is over the L unit vectors pointing at a uniform distribution of points on the complex unit circle,



they cancel each other unless $k = j \pmod{L}$, in which case each term in the sum equals 1.

By the eigenvector condition (A24.51), any matrix function $M = M(r)$ of the translation generator r takes a scalar value on the k th subspace,

$$M(r) \varphi_k = M(\omega^k) \varphi_k, \quad (\text{A24.55})$$

i.e., in the eigenvector basis, M is a diagonal matrix.

The L -dimensional vector $\tilde{\phi}$ of “wavenumbers” (discretized spatial coordinates), or “frequencies,” “eigen-energies” (discretized time evolution steps) $\tilde{\phi}_k$ is the *discrete Fourier transform* of state (vector) ϕ . Hopefully rediscovering it this way helps you a little toward understanding why Fourier transforms are full of $e^{ix \cdot p}$ factors (they are eigenvalues of generators of translations; r for a discrete lattice, $\partial/\partial x$ for continuum), and that they are the natural set of basis functions when a theory is translationally invariant.



example A24.2
p. 1039

A24.4.4 Discrete Fourier transform operator

The $[L \times L]$ matrix $\mathcal{F}_{jk} = L^{-\frac{1}{2}} \omega^{jk}$, $j, k = 0, 1, 2, \dots, L-1$, formed from column eigenvectors (A24.51),

$$\mathcal{F} = \frac{1}{\sqrt{L}} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ 1 & \omega & \omega^2 & \dots & \omega^{L-2} & \omega^{L-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \omega^k & \omega^{2k} & \dots & \omega^{(L-2)k} & \omega^{(L-1)k} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \omega^{L-2} & \omega^{2(L-2)} & \dots & \omega^{(L-2)(L-2)} & \omega^{(L-1)(L-2)} \\ 1 & \omega^{L-1} & \omega^{2(L-1)} & \dots & \omega^{(L-2)(L-1)} & \omega^{(L-1)(L-1)} \end{bmatrix}, \quad (\text{A24.56})$$

is the *discrete Fourier transform operator* (remember, in the discretized world ‘operator’ is a synonym for ‘matrix’). From the orthogonality of eigenvectors (A24.53) it follows that $\mathcal{F}$ is a unitary matrix, with $\det \mathcal{F} = 1$, and

$$\mathcal{F} \mathcal{F}^\dagger = \mathbf{1}. \quad (\text{A24.57})$$

The operator $\mathcal{F}^\dagger$ is thus the *inverse Fourier transform*. The discrete Fourier transform (A24.50) of a state (vector) ϕ is given by

$$\tilde{\phi} = \mathcal{F}^\dagger \phi, \quad (\text{A24.58})$$

i.e., Fourier transformation rearranges components of vector ϕ into averages over all components (A24.50), weighted by complex phases $\exp(i2\pi\ell/L)$ in all possible ways.



example A24.3
p. 1040

The complex function $\tilde{\phi}$ is can sometimes be interpreted as an ‘amplitude function’, with the square of its magnitude ($\tilde{\phi}^\dagger \cdot \tilde{\phi}$) then interpreted as the corresponding ‘total probability’

$$\phi^\dagger \cdot \phi = \tilde{\phi}^\dagger \cdot \tilde{\phi}. \quad (\text{A24.59})$$

The fact that this is the same if evaluated with ϕ or with its Fourier transform $\tilde{\phi}$ is known as the “Parseval’s identity.”

Furthermore, by (A24.55), discrete Fourier transform diagonalizes every translationally invariant matrix function M , i.e., any matrix that commutes with the translation operator, $[r, M] = 0$. To show that, sandwich M with the identity $\mathbf{1} = \mathcal{F} \mathcal{F}^\dagger$:

$$M = \mathbf{1} M \mathbf{1} = \mathcal{F} (\mathcal{F}^\dagger M \mathcal{F}) \mathcal{F}^\dagger = \mathcal{F} \tilde{M} \mathcal{F}^\dagger.$$

The matrix

$$\tilde{M} = \mathcal{F}^\dagger M \mathcal{F} \quad (\text{A24.60})$$

is the Fourier transform of M . The form of any translation-invariant function, such as (A24.59), or the invariant function (A24.39) does not change under $\phi \rightarrow \tilde{\phi}$ transformation, and it does not matter whether we compute in the Fourier space, or in the configuration space that we started out with. For example, the trace of M is the same in either representation

$$\text{tr } M = \text{tr } \mathcal{F} \tilde{M} \mathcal{F}^\dagger = \text{tr } \tilde{M} \mathcal{F}^\dagger \mathcal{F} = \text{tr } \tilde{M},$$

but, if M commutes with the translation operator r , the Fourier transform $\text{tr } \tilde{M}$ is diagonal and trivial to compute. By same reasoning it follows that $\text{tr } M^n = \text{tr } \tilde{M}^n$, and from the $\text{tr } \ln = \ln \text{tr}$ relation that $\det M = \det \tilde{M}$. In fact, any scalar combination of ϕ 's, J 's and couplings, such as the partition function $Z[J]$, has exactly the same form in the configuration and the Fourier space.

Suppose you have two translationally invariant matrices A, B . Evaluating their product AB is a matrix computation. However, evaluating the product in the Fourier space is a simple scalar multiplication of their diagonal elements:

$$(\tilde{A}\tilde{B})_{kk'} = (\mathcal{F}^\dagger A B \mathcal{F})_{kk'} = \tilde{A}_k \tilde{B}_k \delta_{kk'} \quad (\text{A24.61})$$

The continuum Fourier transform version of this relation is called the ‘‘convolution theorem.’’

OK. But what's the payback?

A24.4.5 Lattice Laplacian diagonalized

We can now use the Fourier transform (A24.60) to convert matrix functions of the r matrix into scalars. As M commutes with r , then $(\tilde{M})_{kk'} = \tilde{M}_k \delta_{kk'}$ is a diagonal matrix, where the matrix M acts as a multiplication by the scalar $\tilde{M}_k$ on the k th subspace. For example, for the 1-dimensional version of the lattice Laplacian matrix (A24.32), the eigenvalue condition (A24.51) yields the diagonalized Laplacian in the Fourier space,

$$\begin{aligned} \mathcal{J}_{\mathbb{A}} \varphi_k &= (-\square + \mu^2 \mathbb{I}) \varphi_k = (\mathbf{p}^2 + \mu^2) \varphi_k \\ p &= 2 \sin(k/2), \quad k = \frac{2\pi}{n} m, \quad m = 0, 1, \dots, n-1, \end{aligned} \quad (\text{A24.62})$$

expresses the Fourier-diagonalized lattice Laplacian as the square of the lattice momentum $\mathbf{p}$,

$$(\tilde{\mathcal{J}}_{\mathbb{A}})_{mm'} = (\mathbf{p}^2 + \mu^2) \delta_{mm'}, \quad (\text{A24.63})$$

with n eigenvalues $\Lambda_m = p^2 + \mu^2$ indexed by the integer m . The ‘cord function’ $\text{crd}(\theta) = 2 \sin(\theta/2)$ was used already by Hipparchus cc. 130 BC in the same context, as a discretization of a circle by approximating n arcs by n cords [4, 22].

In the k th subspace the bare propagator is simply a number, and, in contrast to the mess generated by the configuration space inversion (A24.37), there is nothing to inverting M to M^{-1} :

$$(\varphi_{\mathbf{k}}^\dagger \cdot M^{-1} \cdot \varphi_{\mathbf{k}'}) = \frac{\delta_{\mathbf{k}\mathbf{k}'}}{p^2 + \mu^2}, \quad (\text{A24.64})$$

where $\mathbf{k} = (k_1, k_2, \dots, k_d)$ is a d -dimensional vector in the L^d -dimensional dual lattice, i.e., the discretized “momentum” or “frequency” space.

Example: Orbit Jacobian matrix eigenvalues for primitive cell of period 3.

The wave-number range in (A24.62) is $k \in (0, 2\pi/3, 4\pi/3)$. The Laplacian part of the reciprocal lattice orbit Jacobian matrix (A24.63) takes values (up to a sign)

$$p(0) = 0, \quad p(2\pi/3) = \sqrt{3}, \quad .$$

A typical reciprocal lattice site evaluation of p^2 :

$$p_1^2 = p(2\pi/3)^2 = 3.$$

The eigenvalues of the Laplacian p^2 in (A24.63), evaluated on the $V_{\mathbb{A}} = 3$ lattice sites of the reciprocal primitive cell, indexed by integer m , are

$$\begin{bmatrix} p_0^2 & p_1^2 & p_2^2 \end{bmatrix} = \begin{bmatrix} 0 & 3 & 3 \end{bmatrix}, \quad (\text{A24.65})$$

so, for example, the $(\tilde{\mathcal{J}}_{\mathbb{A}})_1$ eigenvalue is

$$\Lambda_2 = 3 + \mu^2,$$

and so on.

Going back to the partition function and sticking in the factors of $\mathbf{1}$ into the bilinear part of the interaction, we replace the spatial source field J by its Fourier transform $\tilde{J}$, and the spatial propagator M by the diagonalized Fourier transformed $\tilde{G}_0$

$$J^\dagger \cdot M \cdot J = J^\dagger \cdot \mathcal{F} (\mathcal{F}^\dagger M \mathcal{F}) \mathcal{F}^\dagger \cdot J = \tilde{J}^\dagger \cdot \tilde{G}_0 \cdot \tilde{J}. \quad (\text{A24.66})$$

A24.5 Continuum field theory

The lattice Laplacian k th Fourier component (A24.99) is

$$\begin{aligned} \tilde{\square}_{kk} &= \frac{2}{a^2} \left(\cos \left(\frac{2\pi}{L} k \right) - 1 \right) \\ &= - \left(\frac{2\pi}{aL} \right)^2 k^2 + \frac{1}{12} \left(\frac{2\pi}{aL} \right)^4 a^2 k^4 - O(k^6). \end{aligned} \quad (\text{A24.67})$$

The quartic term can be neglected for low wave numbers $k \ll L$, i.e., low momenta, $p_\mu = 2\pi k_\mu / LL$, where $aL = LL$ is the lattice size.

In the continuum limit the probability to land in the k th cell is replaced by a probability density, $\phi_k = a^d \phi(x_k) \rightarrow (dx)^d \phi(x)$. After rescaling the wave-number k into momentum p , we obtain the continuum version of the scalar propagator

$$\Delta(x, y) = \int \frac{d^d p}{(2\pi)^d} \frac{e^{ip \cdot (x-y)}}{m^2 + p^2}. \quad (\text{A24.68})$$

A24.6 Diffusion in sawtooth and cat maps

(R. Artuso)



IN THIS SECTION we will deal with the prototype example of chaotic Hamiltonian maps, hyperbolic toral automorphisms. Diffusive properties will arise in considering such maps acting on the cylinder or over $\mathbb{R}^2$, while the dynamics restricted to the fundamental domain involves maps on $\mathbf{T}^2$ (two-dimensional torus). An Anosov map thus corresponds to the action of a matrix in $\text{SL}(2, \mathbb{N})$ with unit determinant and absolute value of the trace bigger than 2.

Maps of this kind are as examples of genuine Hamiltonian chaotic evolution. They admit simple finite Markov partitions, which paves the way to a good symbolic dynamics. Within the framework of Hamiltonian dynamical systems the role of hyperbolic linear automorphisms is analogous to piecewise linear Markov maps: their symbolic dynamics can be encoded in a grammatically simple way, and their linearity leads to uniformity of cycle stabilities.

chapter 28

We will consider the “two-coordinates” representation for them

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = M \begin{bmatrix} x \\ y \end{bmatrix}$$

with

$$M = \begin{bmatrix} 0 & 1 \\ -1 & K+2 \end{bmatrix}$$

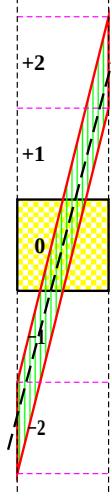
which allows considering their extension on a cylinder phase space $([-1/2, 1/2) \times \mathbb{R})$ in a natural way. So it is natural to study diffusion properties along the y direction.

Though Markov partitions encode the symbolic dynamics in the simplest possible way, they are not well suited to deal with diffusion, as the jumping factor is not related in a simple way to the induced symbol sequence. To this end the following linear code is quite natural: before describing it let us fix the notations: χ will denote the trace of the map ($\chi = K + 2$): the leading eigenvalue will be denoted by $\lambda = (\chi + \sqrt{D})/2$, where $D = \chi^2 - 4$. In principle the code (and the problem of diffusion) can be also considered for real values of K (thus loosing continuity of the torus map when K is not an integer): we will remark in what follows that results which are exact for $K \in \mathbb{N}$ are only approximate for generic K .

The cardinality of the alphabet is determined by the parameter K : the letters are integer numbers, whose absolute values does not exceed $\text{Int}(1 + \chi/2)$ (see figure A24.3 for the case $K = 2$). The code is linear, as, given a bi-infinite sequence $\{x_i\}_{i \in \mathbb{N}}$

$$b_t \stackrel{\text{def}}{=} \left\lfloor (K+2)x_t - x_{t-1} + \frac{1}{2} \right\rfloor, \quad (\text{A24.69})$$

Figure A24.3: The elementary cell for the torus map $[-1/2, 1/2]$ (checkered yellow) together with its image, in green ($K = 2$): symbols refer to the linear code. The dashed line through the origin gives the direction of the unstable manifold. Though hardly understandable from the scale of the picture the unstable manifold is not parallel to the image sides.



[...] denoting the integer part, while the inversion formula (once a bi-infinite symbolic string $\{b_i\}_{i \in \mathbb{N}}$ is given), reads

$$x_t = \frac{1}{\sqrt{D}} \sum_{s \in \mathbb{N}} \lambda^{-|t-s|} b_s, \quad (\text{A24.70})$$

As the x coordinate lives in the interval $[-1/2, 1/2]$, (A24.70) induces a condition of allowed symbol sequences: $\{b_i\}_{i \in \mathbb{N}}$ will be an admissible orbit if

$$\frac{1}{2} \leq \frac{1}{\sqrt{D}} \sum_{s \in \mathbb{N}} \lambda^{-|t-s|} b_s < \frac{1}{2}. \quad (\text{A24.71})$$

By (A24.69) and (A24.70) it is easy to observe that periodic orbits and allowed periodic symbol sequences are in one-to-one correspondence. From (A24.71) we get the condition that a $\{b_i\}_{i=1, \dots, T}$ sequence corresponds to a T -periodic orbit of the map

$$|A_n b_t + A_{n-1}(b_{t+1} + b_{t-1}) + \dots + A_0(b_{t+n} + b_{t-n})| < \frac{B_n}{2} \quad \forall t = 1, \dots, T$$

when $T = 2n + 1$, and

$$|C_n b_t + C_{n-1}(b_{t+1} + b_{t-1}) + \dots + C_0(b_{t+n})| < \frac{D_n}{2} \quad \forall t = 1, \dots, T \quad (\text{A24.72})$$

when $T = 2n$ where

$$\begin{aligned} B_k &= \lambda^k(\lambda - 1) + \lambda^{-k}(\lambda^{-1} - 1) & A_k &= \frac{\lambda^{k+1} + \lambda^{-k}}{\lambda + 1} \\ D_k &= (\lambda^k - \lambda^{-k})(\lambda - \lambda^{-1}) & C_k &= \lambda^k + \lambda^{-k} \end{aligned} \quad (\text{A24.73})$$

exercise A24.3
exercise A24.4

The pruning rules (A24.72) admit a simple geometric interpretation: a lattice point $\mathbf{b} \in \mathbb{N}^T$ identifies a T -periodic point of the map if $\mathbf{b} \in \mathcal{P}_T$ where

$$\mathcal{P}_T \stackrel{\text{def}}{=} \{\mathbf{x} \in \mathbb{R}^T : \begin{cases} |a_1 x_1 + \dots + a_T x_T| < e_T \\ \vdots \\ |a_2 x_1 + \dots + a_1 x_T| < e_T \end{cases}\} \quad (\text{A24.74})$$

and

$$\begin{aligned} a_1 \dots a_T &= A_0 A_1 \dots A_{n-1} A_n A_{n-1} \dots A_0 & e_T &= B_n/2 \\ a_1 \dots a_T &= C_1 \dots C_{n-1} C_n C_{n-1} \dots C_1 C_0 & e_T &= D_n/2 \end{aligned} \quad (\text{A24.75})$$

for $T = 2n+1$ or $T = 2n$, respectively. Thus $\mathcal{P}_T$ is a measure polytope [9], obtained by deforming a T -cube. This is the key issue of this appendix: though the map is endowed with a most remarkable symbolic dynamics, the same is hardly fit to deal with transport properties, as the rectangles that define the partition are not directly connected to translations once the map is unfolded to the cylinder. The partition connected to the linear code (see figure A24.3) on the other side is most natural when dealing with transport, though its not being directly related to invariant manifolds leads to a multitude of pruning rules (which in the present example bear a remarkable geometric interpretation, which is not to be expected as a generic feature).

exercise A24.5

We will denote by $\mathcal{N}_{n,s}$ the number of periodic points of period n with jumping number s . A way to compute D for cat maps is provided by

$$D = \lim_{n \rightarrow \infty} D_n \quad D_n = \frac{1}{n \mathcal{N}_n} \sum_{k=1}^{p(n)} k^2 \mathcal{N}_{n,k} \quad (\text{A24.76})$$

where $\mathcal{N}_n$ is the number of periodic points of period n , $p(n)$ is the highest jumping number of n -periodic orbits and we employed

$$\left| \det \left(\mathbf{1} - \mathbf{J}_x^{(n)} \right) \right| = (\lambda^n - 1)(1 - \lambda^{-n}) = \mathcal{N}_n$$

which is valid for cat maps.

Sums can be converted into integrals by using Poisson summation formula: we define

$$f_T(n) = \begin{cases} (n_1 + \dots + n_T)^2 & n \in \mathcal{P}_T \cap \mathbb{N}^T \\ 0 & \text{otherwise} \end{cases}$$

and

$$\tilde{f}_T(\xi) = \int_{\mathbb{R}^T} dx e^{i(x, \xi)} f_T(x)$$

From Poisson summation formula we have that

$$D_T = \frac{1}{T \mathcal{N}_T} \sum_{n \in \mathbb{N}^T} \tilde{f}_T(2\pi n) \quad (\text{A24.77})$$

The quasilinear estimate for D_T amounts to considering the $n = 0$ contribution to (A24.77):

$$D_T^{(q,l)} = \int_{\mathcal{P}_T} dx (x_1 + x_2 + \dots + x_T)^2 \quad (\text{A24.78})$$

The evaluation of (A24.78) requires introducing a coordinate transformation in symbolic space in which $\mathcal{P}_T$ is transformed in a T -cube. This is equivalent to finding the inverse of the matrix A :

$$A \stackrel{\text{def}}{=} \begin{pmatrix} a_1 & a_2 & \cdots & a_{T-1} & a_T \\ a_T & a_1 & \cdots & a_{T-2} & a_{T-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_3 & a_4 & \cdots & a_1 & a_2 \\ a_2 & a_3 & \cdots & a_T & a_1 \end{pmatrix}. \quad (\text{A24.79})$$

First of all let us observe that A is a circulant matrix, so that its determinant is the product of T factors, each of the form $f(\epsilon_j) = a_1 + \epsilon_j a_2 + \cdots + a_T \epsilon_j^{T-1}$, where ϵ_j is a T th root of unity. By using (A24.73) it is possible to see that

$$f(\epsilon_j) = \begin{cases} \frac{\epsilon_j^{n+1} B_n}{(\lambda \epsilon_j - 1)(1 - \lambda^{-1} \epsilon_j)} & T = 2n + 1 \\ \frac{\epsilon_j^n D_n}{(\lambda \epsilon_j - 1)(1 - \lambda^{-1} \epsilon_j)} & T = 2n \end{cases}$$

so that

$$|\det A| = \frac{(2e_T)^T}{\lambda^T + \lambda^{-T} - 2} \quad (\text{A24.80})$$

By using the results coming from the former exercise we can finally express A^{-1} via

exercise A24.6

$$\tilde{C}A^{-1} = \frac{1}{B_n^T} \begin{pmatrix} \chi & -1 & \cdots & 0 & -1 \\ -1 & \chi & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \chi & -1 \\ -1 & 0 & \cdots & -1 & \chi \end{pmatrix} \quad (\text{A24.81})$$

where

$$\tilde{C} = \begin{pmatrix} \mathbf{0} & \mathbf{1}_{n+1} \\ \mathbf{1}_n & \mathbf{0} \end{pmatrix}.$$

if $T = 2n + 1$ and

$$\tilde{K}A^{-1} = \frac{1}{D_n^T} \begin{pmatrix} \chi & -1 & \cdots & 0 & -1 \\ -1 & \chi & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \chi & -1 \\ -1 & 0 & \cdots & -1 & \chi \end{pmatrix} \quad (\text{A24.82})$$

where

$$\tilde{K} = \begin{pmatrix} \mathbf{0} & \mathbf{1}_n \\ \mathbf{1}_n & \mathbf{0} \end{pmatrix}.$$

if $T = 2n$. As a first check of quasilinear estimates let's compute the volume of $\mathcal{P}_T$:

$$\begin{aligned} \text{Vol}(\mathcal{P}_T) &= \int_{\mathcal{P}_T} dx_1 dx_2 \dots dx_T = \frac{1}{|\det A|} \int_{-e_T}^{e_T} \cdots \int_{e_T}^{e_T} d\xi_1 \dots d\xi_T \\ &= \lambda^T + \lambda^{-T} - 2 \end{aligned} \quad (\text{A24.83})$$

In an analogous way we may compute the quasilinear estimate for $\mathcal{N}_{T,k}$

exercise A24.7

$$\begin{aligned}
 \mathcal{N}_{T,k}^{(q,l)} &= \int_{\mathcal{P}_T} dx_1 \dots dx_T \delta(x_1 + \dots x_T - k) \\
 &= \frac{\lambda^T + \lambda^{-T} - 2}{(2e_T)^T} \int_{-\infty}^{\infty} d\alpha e^{-2\pi i \alpha k} \int_{-e_T}^{e_T} \dots \int_{-e_T}^{e_T} d\xi_1 \dots d\xi_T e^{\frac{2\pi i \alpha \chi}{2e_T} (\xi_1 + \dots + \xi_T)} \\
 &= \frac{2}{\pi \chi} (\lambda^T + \lambda^{-T} - 2) \int_0^{\infty} dy \cos\left(\frac{2qy}{\chi}\right) \left(\frac{\sin y}{y}\right)^T \quad (\text{A24.84})
 \end{aligned}$$

where we have used $x_1 + \dots + x_T = (\chi/(2e_T))(\xi_1 + \dots + \xi_T)$ (cfr. (A24.81),(A24.82)).

We are now ready to evaluate the quasilinear estimate for the diffusion coefficient

$$D_T^{(q,l)} = \frac{1}{\pi \chi T} \int_{-T\chi/2}^{T\chi/2} dz z^2 \int_0^{\infty} dy \cos\left(\frac{2zy}{\chi}\right) \left(\frac{\sin y}{y}\right)^T \quad (\text{A24.85})$$

(where the bounds on the jumping number again come easily from (A24.81),(A24.82)).

By dropping terms vanishing as $T \mapsto \infty$, and using [13]

$$\int_0^{\infty} dx \left(\frac{\sin x}{x}\right)^n \frac{\sin(mx)}{x} = \frac{\pi}{2} \quad m \geq n$$

we can evaluate

$$D^{(q,l)} = \frac{\chi^2}{24} \quad (\text{A24.86})$$

which is the correct result [5] (and again for cat maps (A24.86) is not the quasilinear estimate but the exact value of the diffusion coefficient).

exercise A24.8

Commentary

Remark A24.1. Who has talked about it? (Continued from remark 15.4 Measure polytopes are discussed in ref. [9]. The quasilinear estimate (A24.78) was given in ref. [5]. (A24.78) was evaluated in ref. [2, 20]. Circulant matrix are discussed in ref. [1]. The result (A24.86) agrees with the saw-tooth result of ref. [5]; for the cat maps (A24.86) is the exact value of the diffusion coefficient. This result was obtained, by using periodic orbits also in ref. [11], where Gaussian nature of the diffusion process is explicitly assumed.

Remark A24.2. Discrete Fourier software. [Wolfram Mathematica](#) has an extensive and pedagogical suite of discrete Fourier transform modules.

Remark A24.3. Phase space. The cylinder phase is $[-1/2, 1/2) \times \mathbb{R}$: the map is originally defined on $[-1/2, 1/2)^2$, and is unfolded over the cylinder by symmetry (24.22).

References

- [1] A. C. Aitken, *Determinants & Matrices* (Oliver & Boyd, Edinburgh, 1939).
- [2] R. Artuso and R. Strepparava, “Recycling diffusion in sawtooth and cat maps”, *Phys. Lett. A* **236**, 469–475 (1997).
- [3] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Holt, Rinehart and Winston, 1976).
- [4] E. H. Berger, *Die geographischen Fragmente des Hipparch* (Teubner, 1869).
- [5] J. R. Cary and J. D. Meiss, “Rigorously diffusive deterministic map”, *Phys. Rev. A* **24**, 2664–2668 (1981).
- [6] C.-C. Chen, “Diffusion coefficient of piecewise linear maps”, *Phys. Rev. E* **51**, 2815–2822 (1995).
- [7] D. Cimasoni, “The critical Ising model via Kac-Ward matrices”, *Commun. Math. Phys.* **316**, 99–126 (2012).
- [8] M. M. P. Couchman, D. J. Evans, and J. W. M. Bush, “The stability of a hydrodynamic Bravais lattice”, *Symmetry* **14**, 1524 (2022).
- [9] H. S. M. Coxeter, *Regular Polytopes* (Dover, New York, 1948).
- [10] P. Cvitanović and H. Liang, *A chaotic lattice field theory in two dimensions*, 2025.
- [11] I. Dana, “Hamiltonian transport on unstable periodic orbits”, *Physica D* **39**, 205 (1989).
- [12] C. Godsil and G. F. Royle, *Algebraic Graph Theory* (Springer, New York, 2013).
- [13] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products*, 8th ed. (Elsevier LTD, Oxford, New York, 2014).
- [14] G. W. Hill, “On the part of the motion of the lunar perigee which is a function of the mean motions of the sun and moon”, *Acta Math.* **8**, 1–36 (1886).
- [15] R. Klages, *Deterministic Diffusion in One-dimensional Chaotic Dynamical Systems* (Wissenschaft and Technik-Verlag, Berlin, 1996).
- [16] R. Klages and J. R. Dorfman, “Simple maps with fractal diffusion coefficients”, *Phys. Rev. Lett.* **74**, 387–390 (1995).
- [17] R. Klages and J. R. Dorfman, “Dynamical crossover in deterministic diffusion”, *Phys. Rev. E* **55**, R1247–R1250 (1997).
- [18] D. A. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding* (Cambridge Univ. Press, Cambridge UK, 1995).
- [19] M. Pollicott, *Dynamical zeta functions*, in *Smooth Ergodic Theory and Its Applications*, Vol. 69, edited by A. Katok, R. de la Llave, Y. Pesin, and H. Weiss (2001), pp. 409–428.
- [20] R. Strepparava, Laurea thesis, MA thesis (Università degli Studi di Milano, 1995).

- [21] H.-C. Tseng, H.-J. Chen, P.-C. Li, W.-Y. Lai, C.-H. Chou, and H.-W. Che, “Some exact results for the diffusion coefficients of maps with pruning cycles”, *Phys. Lett. A* **195**, 74–80 (1994).
- [22] Wikipedia contributors, [Chord \(geometry\) — Wikipedia, The Free Encyclopedia](#), 2023.

A24.7 Examples

Example A24.1. A 2-point lattice diagonalized.

The action of the shift operator d (A24.15) on a 2-point lattice $\phi = (\phi_0, \phi_1)$ is to permute the two lattice sites

$$d = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

As exchange repeated twice brings us back to the original state, $d^2 = \mathbf{1}$, the characteristic polynomial of d is

$$(d + 1)(d - 1) = 0,$$

with eigenvalues $\omega_0 = 1, \omega_1 = -1$. The symmetrization, antisymmetrization projection operators are

$$P_0 = \frac{d - \omega_1 \mathbf{1}}{\omega_0 - \omega_1} = \frac{1}{2}(\mathbf{1} + d) = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad (\text{A24.87})$$

$$P_1 = \frac{d - \mathbf{1}}{-1 - 1} = \frac{1}{2}(\mathbf{1} - d) = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}. \quad (\text{A24.88})$$

Noting that $P_0 + P_1 = \mathbf{1}$, we can project a lattice state ϕ onto the two normalized eigenvectors of d :

$$\begin{aligned} \phi &= \mathbf{1} \phi = P_0 \cdot \phi + P_1 \cdot \phi, \\ \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} &= \frac{(\phi_0 + \phi_1)}{\sqrt{2}} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{(\phi_0 - \phi_1)}{\sqrt{2}} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \end{aligned} \quad (\text{A24.89})$$

$$= \tilde{\phi}_0 \varphi_0 + \tilde{\phi}_1 \varphi_1. \quad (\text{A24.90})$$

As $P_0 P_1 = 0$, the symmetric and the antisymmetric states transform separately under any linear transformation constructed from d and its powers.

In this way the characteristic equation $d^2 = \mathbf{1}$ enables us to reduce the 2-dimensional lattice state to two 1-dimensional ones, on which the value of the shift operator d is a number, $\omega_j \in \{1, -1\}$, and the normalized eigenvectors are $\varphi_0 = \frac{1}{\sqrt{2}}(1, 1)$, $\varphi_1 = \frac{1}{\sqrt{2}}(1, -1)$. As we shall now see, $(\tilde{\phi}_0, \tilde{\phi}_1)$ is the 2-site periodic lattice discrete Fourier transform of the field (ϕ_1, ϕ_2) .

[return to p. 1025](#)

Example A24.2. Cyclic group projection operators AKA discrete Fourier transform

(It's OK to skip this example on the first reading - the explicit Fourier eigenvectors and eigenvalues (A24.51) are all that we need to carry out discrete Fourier transforms.)



Consider a cyclic group

$$C_N = \{e, g, g^2, \dots, g^{N-1}\}, \quad g^N = e.$$

If $M = D(g)$ is a $[d \times d]$ matrix representation of the one-step shift g , it must satisfy $M^N - \mathbf{1} = 0$, with eigenvalues given by the zeros of the characteristic polynomial

$$G(x) = x^N - 1 = (x - \lambda_0)(x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_{N-1}). \quad (\text{A24.91})$$

For the cyclic group the N distinct eigenvalues are the N th roots of unity $\lambda_n = \omega^n$, $\omega = \exp(i 2\pi/N)$, $n = 0, \dots, N-1$.

In the projection operator formulation (A10.20), they split the d -dimensional space into d/N -dimensional subspaces by means of projection operators

$$P_n = \prod_{m \neq n} \frac{M - \omega^m I}{\omega^n - \omega^m} = \frac{1}{\prod_{m=1}^{N-1} (1 - \omega^m)} \prod_{m=1}^{N-1} (\omega^{-n} M - \omega^m I), \quad (\text{A24.92})$$

where we have multiplied all denominators and numerators by ω^{-n} .

The denominator is a polynomial of form $G(x)/(x - \lambda_0)$, with the zeroth root $(x - \omega^0) = (x - 1)$ quotiented out from the characteristic polynomial,

$$\frac{x^N - 1}{x - 1} = (x - \omega)(x - \omega^2) \cdots (x - \omega^{N-1}).$$

Consider a sum of the first N terms of a geometric series, multiplied by $(x - 1)/(x - 1)$:

$$1 + x + \cdots + x^{N-1} = \sum_{m=0}^{N-1} x^m = \frac{1}{x - 1} \sum_{m=0}^{N-1} (x - 1) x^m = \frac{x^N - 1}{x - 1}. \quad (\text{A24.93})$$

So, the products in (A24.92) can be written as sums

$$(x - \omega)(x - \omega^2) \cdots (x - \omega^{N-1}) = 1 + x + \cdots + x^{N-1}. \quad (\text{A24.94})$$

The P_n projection operator (A24.92) denominator is evaluated by substituting $x \rightarrow 1$ into (A24.94); that adds up to N . The numerator is evaluated by substituting $x \rightarrow \omega^{-n} M$. We obtain the projection operator as a discrete Fourier weighted sum of matrices M^m ,

$$P_n = \frac{1}{N} \sum_{m=0}^{N-1} e^{-i \frac{2\pi}{N} nm} M^m, \quad (\text{A24.95})$$

instead of the product form (A24.92).

This is the simplest example of the key group theory tool, the projection operator expressed as a sum over characters,

$$P_n = \frac{1}{|G|} \sum_{g \in G} \tilde{\chi}_n(g) D(g).$$

As C_N irreps are all 1-dimensional, for the discrete Fourier transform all characters are simply $\tilde{\chi}_n(g^m) = \omega^{-nm}$, the N th complex roots of unity.

(B. Gutkin and P. Cvitanović)

[return to p. 1028](#)

Example A24.3. ‘Configuration-momentum’ Fourier space duality.

What does a projection on the k th Fourier subspace mean? The discrete Fourier transform (A24.95) of a state (vector) ϕ rearranges components of vector ϕ into averages over all its components, weighted by complex phases $\exp(i 2\pi \ell/L)$ in all possible ways.

Consider first the projection on the 0th Fourier mode

$$P_0 = \frac{1}{L} \sum_{m=0}^{L-1} d^m.$$

Applied to a lattice state $\phi = (\phi_1, \phi_2, \dots, \phi_L)$, the shift matrix d translates the state by one site, $d\phi = (\phi_2, \phi_3, \dots, \phi_L, \phi_1)$, and so on for all powers d^m . The result is the space average (here correctly normalized, so that $\langle 1 \rangle = 1$) over all values of the periodic lattice field ϕ_m ,

$$\frac{1}{\sqrt{L}}\tilde{\phi}_0 = \frac{1}{L} \sum_{\ell=0}^{L-1} \phi_\ell = \langle \phi \rangle,$$

see (A24.43) and (A24.54). Every finite discrete group has such fully-symmetric representation, and in statistical mechanics and quantum mechanics this is often the most important state (the ‘ground’ state).

$\tilde{\phi}_1$ is the average weighted by one oscillation over the L -periodic lattice, and $\tilde{\phi}_k$, the projection of the L -dimensional state (i.e., vector) ϕ on the k th subspace

$$\tilde{\phi}_k = P_k \cdot \phi = \frac{1}{\sqrt{L}} \sum_{\ell=0}^{L-1} \omega^{k\ell} \phi_\ell, \quad (\text{A24.96})$$

is the average weighted by complex rotating phase ω^{km} which advances by ω^k in every step, and pulls out oscillating feature $\tilde{\phi}_k$ out of the field ϕ . For large L , modes $\tilde{\phi}_k$ with $k \ll L$ (or $(L - k) \ll L$, that is just a counter-rotation)) are called *hydrodynamic modes*, corresponding to “configuration” lattice fields ϕ which vary slowly and smoothly over many lattice spacings. Modes with $k \simeq L/2$ are suspect, they are lattice discretization artifacts.

If the lattice state ϕ is localized, its Fourier transform will be global, and vice versa for a localized Fourier state $\tilde{\phi}$. For example, if the field ϕ is concentrated on the first site, $\phi_0 = 1$, rest zero, its Fourier transform will be uniformly distributed over all Fourier modes, $\tilde{\phi}_k = 1/\sqrt{L}$.

[return to p. 1029](#)

Exercises

A24.1. Laplacian is a non-local operator.

While the Laplacian is a simple tri-diagonal difference operator (A24.31), its inverse (the “free” propagator of statistical mechanics and quantum field theory) is a messier object. A way to compute is to start expanding propagator as a power series in the Laplacian

$$\frac{1}{m^2 \mathbf{1} - \Delta} = \frac{1}{m^2} \sum_{n=0}^{\infty} \frac{1}{m^{2n}} \Delta^n. \quad (\text{A24.97})$$

As Δ is a finite matrix, the expansion is convergent for sufficiently large m^2 . To get a feeling for what is involved in evaluating such series, show that Δ^2 is:

$$\Delta^2 = \frac{1}{a^4} \begin{bmatrix} 6 & -4 & 1 & & 1 & -4 \\ -4 & 6 & -4 & 1 & & \\ 1 & -4 & 6 & -4 & 1 & \\ & 1 & -4 & \ddots & & \\ & & & & 6 & -4 \\ -4 & 1 & & & 1 & -4 & 6 \end{bmatrix}. \quad (\text{A24.98})$$

What $\Delta^3, \Delta^4, \dots$ contributions look like is now clear; as we include higher and higher powers of the Laplacian, the propagator matrix fills up; while the *inverse* propagator is differential operator connecting only the nearest neighbors, the propagator is integral operator, connecting every lattice site to any other lattice site.

This matrix can be evaluated as is, on the lattice, and sometime it is evaluated this way, but in case at hand a wonderful simplification follows from the observation that the lattice action is translationally invariant, exercise A24.2.

A24.2. Lattice Laplacian diagonalized. Insert the identity $\sum P^{(k)} = \mathbf{1}$ wherever you profitably can, and use the eigenvalue equation (A24.51) to convert shift σ matrices into scalars. If M commutes with σ , then $(\varphi_k^\dagger \cdot M \cdot \varphi_{k'}) = \tilde{M}^{(k)} \delta_{kk'}$, and the matrix M acts as a multiplication by the scalar $\tilde{M}^{(k)}$ on the k th subspace. Show that for the 1-dimensional version of the lattice Laplacian (A24.31) the projection on the k th subspace is

$$(\varphi_k^\dagger \cdot \Delta \cdot \varphi_{k'}) = \frac{2}{a^2} \left(\cos \left(\frac{2\pi}{N} k \right) - 1 \right) \delta_{kk'}. \quad (\text{A24.99})$$

In the k th subspace the propagator is simply a number, and, in contrast to the mess generated by (A24.97), there is nothing to evaluating:

$$\varphi_k^\dagger \cdot \frac{1}{m^2 \mathbf{1} - \Delta} \cdot \varphi_{k'} = \frac{\delta_{kk'}}{m^2 - \frac{2}{(ma)^2} (\cos 2\pi k/N - 1)}, \quad (\text{A24.100})$$

where k is a site in the N -dimensional dual lattice, and $a = L/N$ is the lattice spacing.

A24.3. Recursion relations. Verify that the following recursion relations are satisfied

$$u_{k+2} = \chi u_{k+1} - u_k$$

where $u_k = A_k, B_k, C_k, D_k$.

A24.4. **Arnol'd cat map.** Show that for $\chi = 3$, $A_k = F_{2k+1}$, $B_k = L_{2k+1}$, $C_k = L_{2k}$ and $D_k = 5F_{2k}$, where F_n and L_n are the Fibonacci and Lucas numbers..

A24.5. **Pruning rules for substrings of length 2.** Take $K = 8$ and draw the region determined by (A24.72).

A24.6. **Diagonalization of A.** Show that A can be diagonalized by considering the auxiliary matrix U

$$U \stackrel{\text{def}}{=} \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ \epsilon_0 & \epsilon_1 & \cdots & \epsilon_{T-2} & \epsilon_{T-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \epsilon_0^{T-2} & \epsilon_1^{T-2} & \cdots & \epsilon_{T-2}^{T-2} & \epsilon_{T-1}^{T-2} \\ \epsilon_0^{T-1} & \epsilon_1^{T-1} & \cdots & \epsilon_{T-2}^{T-1} & \epsilon_{T-1}^{T-1} \end{pmatrix}.$$

In fact $U^{-1}AU$ is a diagonal matrix (the diagonal elements coinciding with $f(\epsilon_j)$).

A24.7. **Periodic points of cat maps.** Verify that (A24.83) is exactly the number of T -periodic points of the map when K is an integer.

A24.8. **Probability distribution.** Higher order moments can be computed easily for integer K (or generic K within the quasilinear approximation), by generalizations of (A24.85): show that the results prove that, given a period T , the distribution of periodic orbits with respect to their jumping number is asymptotically Gaussian, with parameter $D^{(q,l)}$.

A24.9. **Deterministic diffusion, zig-zag map.**

To illustrate the main idea of chapter 24, tracking of a globally diffusing orbit by the associated confined orbit restricted to the fundamental cell, we consider a class of simple 1-dimensional dynamical systems, chains of piecewise linear maps, where all transport coefficients can be evaluated analytically. The translational symmetry (24.22) relates the unbounded dynamics on the real line to the dynamics restricted to a “fundamental cell” - in the present example the unit interval curled up into a circle. An example of such map is the sawtooth map

$$\hat{f}(x) = \begin{cases} \Lambda x & x \in [0, 1/4 + 1/4\Lambda] \\ -\Lambda x + (\Lambda + 1)/2 & x \in [1/4 + 1/4\Lambda, 3/4 - 1/4\Lambda] \\ \Lambda x + (1 - \Lambda) & x \in [3/4 - 1/4\Lambda, 1] \end{cases}. \quad (\text{A24.101})$$

The corresponding circle map $f(x)$ is obtained by modulo the integer part. The elementary cell map $f(x)$ is sketched in figure A24.4. The map has the symmetry property

$$\hat{f}(\hat{x}) = -\hat{f}(-\hat{x}), \quad (\text{A24.102})$$

so that the dynamics has no drift, and all odd derivatives of the generating function (24.4) with respect to β evaluated at $\beta = 0$ vanish.

The cycle weights are given by

$$t_p = z^{n_p} \frac{e^{\beta \hat{n}_p}}{|\Lambda_p|}. \quad (\text{A24.103})$$

The diffusion constant formula for 1-dimensional maps is

$$D = \frac{1}{2} \frac{\langle \hat{n}^2 \rangle_\zeta}{\langle n \rangle_\zeta} \quad (\text{A24.104})$$

where the “mean cycle time” is given by

$$\langle n \rangle_\zeta = z \frac{\partial}{\partial z} \frac{1}{\zeta(0, z)} \Big|_{z=1} = - \sum' (-1)^k \frac{n_{p_1} + \cdots + n_{p_k}}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}, \quad (\text{A24.105})$$

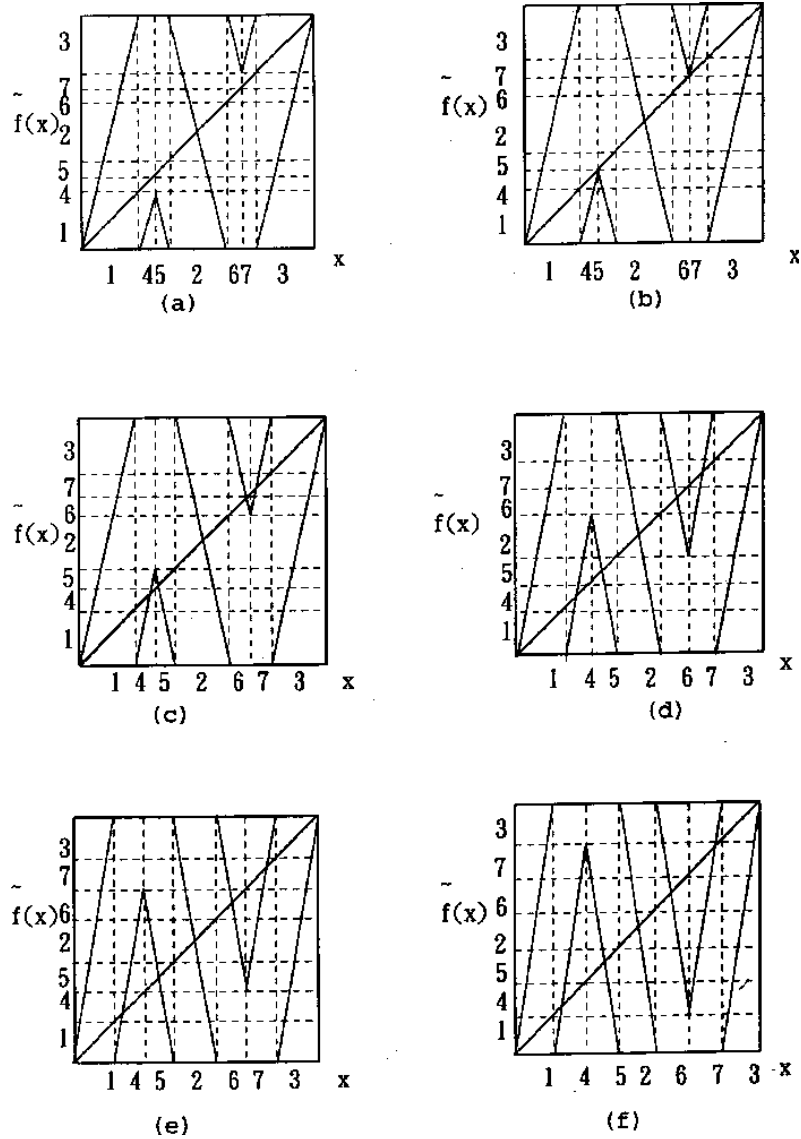


Figure A24.4: (a)-(f) The sawtooth map (A24.101) for the 6 values of parameter a for which the folding point of the map aligns with the endpoint of one of the 7 intervals and yields a finite Markov partition (from ref. [21]). The corresponding transition graphs are given in figure A24.5.

the mean cycle displacement squared by

$$\langle \hat{n}^2 \rangle_\zeta = \left. \frac{\partial^2}{\partial \beta^2} \frac{1}{\zeta(\beta, 1)} \right|_{\beta=0} = - \sum' (-1)^k \frac{(\hat{n}_{p_1} + \dots + \hat{n}_{p_k})^2}{|\Lambda_{p_1} \dots \Lambda_{p_k}|}, \quad (\text{A24.106})$$

and the sum is over all distinct non-repeating combinations of prime cycles. Most of results expected in this projects require no more than pencil and paper computations.

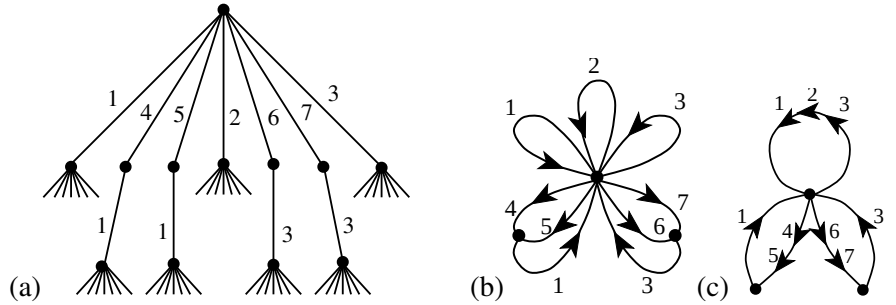
Implementing the symmetry factorization (24.19) is convenient, but not essential for this project, so if you find example 25.9 too long a read, skip the symmetrization.

A24.10. The full shift sawtooth map. Take the map (A24.101) and extend it to the real line. As in example of figure 24.4, denote by a the critical value of the map (the maximum height in the unit cell)

$$a = \hat{f}\left(\frac{1}{4} + \frac{1}{4\Lambda}\right) = \frac{\Lambda + 1}{4}. \quad (\text{A24.107})$$

Describe the symbolic dynamics that you obtain when a is an integer, and derive the formula

Figure A24.5: (a) The sawtooth map (A24.101) partition tree for figure A24.4(a); while intervals $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ map onto the whole unit interval, $f(\mathcal{M}_1) = f(\mathcal{M}_2) = f(\mathcal{M}_3) = \mathcal{M}$, intervals $\mathcal{M}_4, \mathcal{M}_5$ map onto $\mathcal{M}_1$ only, $f(\mathcal{M}_4) = f(\mathcal{M}_5) = \mathcal{M}_1$, and similarly for intervals $\mathcal{M}_6, \mathcal{M}_7$. An initial point starting out in the interval $\mathcal{M}_1, \mathcal{M}_2$ or $\mathcal{M}_3$ can land anywhere on the unit interval, so the subtrees originating from the corresponding nodes on the partition tree are similar to the whole tree and can be identified as (e. g. figure 17.6) yielding (b) the transition graph for the Markov partition of figure A24.4(a) and (c) the transition graph in the compact notation of (24.32).



for the diffusion constant:

$$D = \frac{(\Lambda^2 - 1)(\Lambda - 3)}{96\Lambda} \quad \text{for } \Lambda = 4a - 1, \quad a \in \mathbb{Z}. \quad (\text{A24.108})$$

If you are going strong, derive also the formula for the half-integer $a = (2k + 1)/2$, $\Lambda = 4a + 1$ case and email it to predrag@nbi.dk. You will need to partition $\mathcal{M}_2$ into the left and right half, $\mathcal{M}_2 = \mathcal{M}_8 \cup \mathcal{M}_9$, as in the derivation of (24.29). See exercise 24.1.

A24.11. Sawtooth map subshifts of finite type. We now work out an example when the partition is Markov, although the slope is not an integer number. The key step is that of having a partition where intervals are mapped *onto* unions of intervals. Consider for example the case in which $\Lambda = 4a - 1$, where $1 \leq a \leq 2$. A first partition is constructed from seven intervals, which we label $\{\mathcal{M}_1, \mathcal{M}_4, \mathcal{M}_5, \mathcal{M}_2, \mathcal{M}_6, \mathcal{M}_7, \mathcal{M}_3\}$, with the alphabet ordered as the intervals are laid out along the unit interval. In general the critical value a will not correspond to an interval border, but now we choose a such that the critical point is mapped onto the right border of $\mathcal{M}_1$, as in figure A24.4(a). The critical value of $f()$ is $f(\frac{\Lambda+1}{4\Lambda}) = a - 1 = (\Lambda - 3)/4$. Equating this with the right border of $\mathcal{M}_1$, $x = 1/\Lambda$, we obtain a quadratic equation with the expanding solution $\Lambda = 4$. We have that $f(\mathcal{M}_4) = f(\mathcal{M}_5) = \mathcal{M}_1$, so the transition matrix (17.1) is given by

$$\phi' = T\phi = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_4 \\ \phi_5 \\ \phi_2 \\ \phi_6 \\ \phi_7 \\ \phi_3 \end{bmatrix} \quad (\text{A24.109})$$

and the dynamics is unrestricted in the alphabet

$$\{1, \underline{41}, \underline{51}, 2, \underline{63}, \underline{73}, 3, \}.$$

One could diagonalize (A24.109) on the computer, but, as we saw in chapter 17, the transition graph figure A24.5 (b) corresponding to figure A24.4(a) offers more insight into the dynamics. The dynamical zeta function

$$\begin{aligned} 1/\zeta &= 1 - (t_1 + t_2 + t_3) - 2(t_{14} + t_{37}) \\ 1/\zeta &= 1 - 3\frac{z}{\Lambda} - 4 \cosh \beta \frac{z^2}{\Lambda^2}. \end{aligned} \quad (\text{A24.110})$$

follows from the loop expansion (18.13) of sect. 18.3.

The material flow conservation sect. 23.4 and the symmetry factorization (24.19) yield

$$0 = \frac{1}{\zeta(0, 1)} = \left(1 + \frac{1}{\Lambda}\right) \left(1 - \frac{4}{\Lambda}\right)$$

figure A24.4	Λ	D
	3	0
(a)	4	$\frac{1}{10}$
(b)	$\sqrt{5} + 2$	$\frac{1}{2\sqrt{5}}$
(c)	$\frac{1}{2}(\sqrt{17} + 5)$	$\frac{2}{\sqrt{17}}$
(c')	5	$\frac{2}{5}$
(d)	$\frac{1}{2}(\sqrt{33} + 5)$	$\frac{1}{8} + \frac{5}{88}\sqrt{33}$
(e)	$2\sqrt{2} + 3$	$\frac{1}{2\sqrt{2}}$
(f)	$\frac{1}{2}(\sqrt{33} + 7)$	$\frac{1}{4} + \frac{1}{4\sqrt{33}}$
	7	$\frac{2}{7}$

Table A24.1: The diffusion constant as function of the slope Λ for the $a = 1, 2$ values of (A24.108) and the 6 Markov partitions of figure A24.4

which indeed is satisfied by the given value of Λ . Conversely, we can use the desired Markov partition topology to write down the corresponding dynamical zeta function, and use the $1/\zeta(0, 1) = 0$ condition to fix Λ . For more complicated transition matrices the factorization (24.19) is very helpful in reducing the order of the polynomial condition that fixes Λ .

The diffusion constant follows from (24.20) and (A24.104)

$$\langle n \rangle_\zeta = - \left(1 + \frac{1}{\Lambda} \right) \left(-\frac{4}{\Lambda} \right), \quad \langle \hat{n}^2 \rangle_\zeta = \frac{4}{\Lambda^2}$$

$$D = \frac{1}{2} \frac{1}{\Lambda + 1} = \frac{1}{10}$$

Think up other non-integer values of the parameter for which the symbolic dynamics is given in terms of Markov partitions: in particular consider the cases illustrated in figure A24.4 and determine for what value of the parameter a each of them is realized. Work out the transition graph, symmetrization factorization and the diffusion constant, and check the material flow conservation for each case. Derive the diffusion constants listed in table A24.1. It is not clear why the final answers tend to be so simple. Numerically, the case of figure A24.4 (c) appears to yield the maximal diffusion constant. Does it? Is there an argument that it should be so?

The seven cases considered here (see table A24.1, figure A24.4 and (A24.108)) are the 7 simplest complete Markov partitions, the criterion being that the critical points map onto partition boundary points. This is, for example, what happens for unimodal tent map; if the critical point is preperiodic to an unstable cycle, the grammar is complete. The simplest example is the case in which the tent map critical point is preperiodic to a unimodal map 3-cycle, in which case the grammar is of golden mean type, with `_00_` substring prohibited (see figure 17.6). In case at hand, the “critical” point is the junction of branches 4 and 5 (symmetry automatically takes care of the other critical point, at the junction of branches 6 and 7), and for the cases considered the critical point maps into the endpoint of each of the seven branches.

One can fill out parameter a axis arbitrarily densely with such points - each of the 7 primary intervals can be subdivided into 7 intervals obtained by 2-nd iterate of the map, and for the critical point mapping into any of those in 2 steps the grammar (and the corresponding cycle expansion) is finite, and so on.

A24.12. Sawtooth map diffusion coefficient, numerically. (optional:)

Attempt a numerical evaluation of

$$D = \frac{1}{2} \lim_{n \rightarrow \infty} \frac{1}{n} \langle \hat{x}_n^2 \rangle. \quad (\text{A24.111})$$

Study the convergence by comparing your numerical results to the exact answers derived above. Is it better to use few initial $\hat{x}$ and average for long times, or to use many initial $\hat{x}$ for

shorter times? Or should one fit the distribution of $\hat{x}^2$ with a Gaussian and get the D this way? Try to plot dependence of D on Λ ; perhaps blow up a small region to show that the dependence of D on the parameter Λ is fractal. Compare with figure 24.6 and figures in refs. [6, 15–17, 21].

A24.13. **Sawtooth D is a nonuniform function of the parameters.** (optional:)

The dependence of D on the map parameter Λ is rather unexpected - even though for larger Λ more points are mapped outside the unit cell in one iteration, the diffusion constant does not necessarily grow. An interpretation of this lack of monotonicity would be interesting.

You can also try applying periodic orbit theory to the sawtooth map (A24.101) for a random “generic” value of the parameter Λ , for example $\Lambda = 6$. The idea is to bracket this value of Λ by the nearby ones, for which higher and higher iterates of the critical value $a = (\Lambda + 1)/4$ fall onto the partition boundaries, compute the exact diffusion constant for each such approximate Markov partition, and study their convergence toward the value of D for $\Lambda = 6$. Judging how difficult such problem is already for a tent map (see sect. 18.5 and appendix A18.1), this is too ambitious for a week-long exam.

A24.14. **Deterministic diffusion, sawtooth map.**

To illustrate the main idea of chapter 24, tracking of a globally diffusing orbit by the associated confined orbit restricted to the fundamental cell, we consider in more detail the class of simple 1-dimensional dynamical systems, chains of piecewise linear maps (24.21). The translational symmetry (24.22) relates the unbounded dynamics on the real line to the dynamics restricted to a “fundamental cell” - in the present example the unit interval curled up into a circle. The corresponding circle map $f(x)$ is obtained by modulo the integer part. The elementary cell map $f(x)$ is sketched in figure 24.4. The map has the symmetry property

$$\hat{f}(\hat{x}) = -\hat{f}(-\hat{x}), \quad (\text{A24.112})$$

so that the dynamics has no drift, and all odd derivatives of the generating function (24.4) with respect to β evaluated at $\beta = 0$ vanish.

The cycle weights are given by

$$t_p = z^{n_p} \frac{e^{\beta \hat{n}_p}}{|\Lambda_p|}. \quad (\text{A24.113})$$

The diffusion constant formula for 1-dimensional maps is

$$D = \frac{1}{2} \frac{\langle \hat{n}^2 \rangle_\zeta}{\langle n \rangle_\zeta} \quad (\text{A24.114})$$

where the “mean cycle time” is given by

$$\langle n \rangle_\zeta = z \frac{\partial}{\partial z} \frac{1}{\zeta(0, z)} \Big|_{z=1} = - \sum' (-1)^k \frac{n_{p_1} + \cdots + n_{p_k}}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}, \quad (\text{A24.115})$$

the mean cycle displacement squared by

$$\langle \hat{n}^2 \rangle_\zeta = \frac{\partial^2}{\partial \beta^2} \frac{1}{\zeta(\beta, 1)} \Big|_{\beta=0} = - \sum' (-1)^k \frac{(\hat{n}_{p_1} + \cdots + \hat{n}_{p_k})^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}, \quad (\text{A24.116})$$

and the sum is over all distinct non-repeating combinations of prime cycles. Most of results expected in this projects require no more than pencil and paper computations.

Appendix A25

Discrete symmetry factorization

A25.1 D_4 factorization

If an n -disk arrangement has C_n symmetry, and the disk visitation sequence is given by disk labels $\{\epsilon_1 \epsilon_2 \epsilon_3 \dots\}$, only the relative increments $\rho_i = \epsilon_{i+1} - \epsilon_i \bmod n$ matter. Symmetries under reflections across axes increase the group to D_n and add relations between symbols: $\{\epsilon_i\}$ and $\{n - \epsilon_i\}$ differ only by a reflection. As a consequence of this reflection increments become decrements until the next reflection and vice versa. Consider four equal disks placed on the vertices of a square (figure A25.1). The symmetry group consists of the identity e , the two reflections s, s_2 across x, y axes, the two diagonal reflections s_1, s_3 , and the three rotations r, r_2 and r_3 by angles $\pi/2, \pi$ and $3\pi/2$. We start by exploiting the C_4 subgroup symmetry in order to replace the absolute labels $\epsilon_i \in \{1, 2, 3, 4\}$ by relative increments $\rho_i \in \{1, 2, 3\}$. By reflection across diagonals, an increment by 3 is equivalent to an increment by 1 and a reflection; this new symbol will be called $\underline{1}$. Our convention will be to first perform the increment and then to change the orientation due to the reflection. As an example, consider the fundamental domain cycle 112 . Taking the disk $1 \rightarrow$ disk 2 segment as the starting segment, this symbol string is mapped into the disk visitation sequence $1_{+1}2_{+1}3_{+2}1 \dots = \overline{123}$, where the subscript indicates the increments (or decrements) between neighboring symbols; the period of the cycle $\overline{112}$ is thus 3 in both the fundamental domain and the full space. Similarly, the cycle $\underline{112}$ will be mapped into $1_{+1}2_{-1}1_{-2}3_{-1}2_{+1}3_{+2}1 = \overline{121323}$ (note that the fundamental domain symbol $\underline{1}$ corresponds to a flip in orientation after the second and fifth symbols); this time the period in the full space is twice that of the fundamental domain. In particular, the fundamental domain fixed points correspond to the following 4-disk cycles:

4-disk		reduced
12	$\leftrightarrow$	$\frac{1}{1}$
1234	$\leftrightarrow$	$\frac{1}{1}$
13	$\leftrightarrow$	2

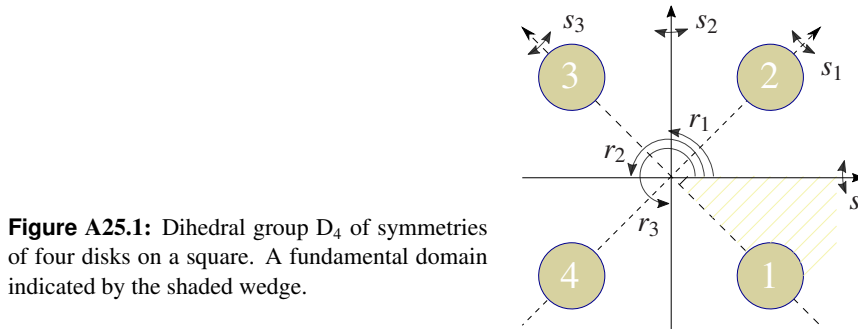


Figure A25.1: Dihedral group D_4 of symmetries of four disks on a square. A fundamental domain indicated by the shaded wedge.

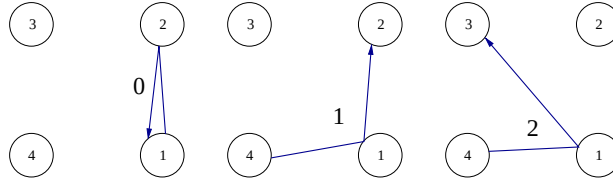


Figure A25.2: Reduced, fundamental domain symbolic dynamics for four disks on a square.

Conversions for all periodic orbits of reduced symbol period less than 5 are listed in table [A25.1](#).

[section 38.1.1](#)

This symbolic dynamics is closely related to the group-theoretic structure of the dynamics: the global 4-disk trajectory can be generated by mapping the fundamental domain trajectories onto the full 4-disk space by the accumulated product of the D_4 group elements $g_1 = r$, $g_2 = r_2$, $g_{\perp} = sr = s_1$, where r is a quarter rotation by $\pi/2$. In the $\overline{112}$ example worked out above, this yields $g_{\perp 12} = g_2 g_1 g_{\perp} = r_2 r s = s_3$, listed in the last column of table [A25.1](#). Our convention is to multiply group elements in the reverse order with respect to the symbol sequence. We need these group elements for our next step, the dynamical zeta function factorizations.

The D_4 group has four 1-dimensional representations, either symmetric (A_1) or antisymmetric (A_2) under both types of reflections, or symmetric under one and antisymmetric under the other (B_1 , B_2), and a 2-dimensional representation E . Substituting the D_4 characters

D_4	A_1	A_2	B_1	B_2	E
e	1	1	1	1	2
r_2	1	1	1	1	-2
r, r_3	1	1	-1	-1	0
s, s_2	1	-1	1	-1	0
s_1, s_3	1	-1	-1	1	0

into [\(25.20\)](#) we obtain:

Table A25.1: D_4 correspondence between the ternary fundamental domain prime cycles $\tilde{p}$ and the full 4-disk $\{1,2,3,4\}$ labeled cycles p , together with the D_4 transformation that maps the end point of the $\tilde{p}$ cycle into an irreducible segment of the p cycle. For typographical convenience, the symbol $\underline{1}$ of sect. A25.1 has been replaced by 0, so that the ternary alphabet is $\{0, 1, 2\}$. The degeneracy of the p cycle is $m_p = 8n_{\tilde{p}}/n_p$. Orbit $\bar{2}$ is the sole boundary orbit, invariant both under a rotation by π and a reflection across a diagonal. The two pairs of cycles marked by (a) and (b) are related by time reversal, but cannot be mapped into each other by D_4 transformations.

$\tilde{p}$	p	$\mathbf{h}_{\tilde{p}}$	$\tilde{p}$	p	$\mathbf{h}_{\tilde{p}}$
0	1 2	σ_x	0001	1212 1414	σ_{24}
1	1 2 3 4	r	0002	1212 4343	σ_y
2	1 3	r_2, σ_{13}	0011	1212 3434	r_2
01	12 14	σ_{24}	0012	1212 4141 3434 2323	r_3
02	12 43	σ_y	0021 (a)	1213 4142 3431 2324	r_3
12	12 41 34 23	r_3	0022	1213	e
001	121 232 343 414	r	0102 (a)	1214 2321 3432 4143	r
002	121 343	r_2	0111	1214 3234	σ_{13}
011	121 434	σ_y	0112 (b)	1214 2123	σ_x
012	121 323	σ_{13}	0121 (b)	1213 2124	σ_x
021	124 324	σ_{13}	0122	1213 1413	σ_{24}
022	124 213	σ_x	0211	1243 2134	σ_x
112	123	e	0212	1243 1423	σ_{24}
122	124 231 342 413	r	0221	1242 1424	σ_{24}
			0222	1242 4313	σ_y
			1112	1234 2341 3412 4123	r
			1122	1231 3413	r_2
			1222	1242 4131 3424 2313	r_3

$$\begin{array}{lcl}
h_{\tilde{p}} & & A_1 \quad A_2 \quad B_1 \quad B_2 \quad E \\
e: & (1 - t_{\tilde{p}})^8 = & (1 - t_{\tilde{p}}) \quad (1 - t_{\tilde{p}}) \quad (1 - t_{\tilde{p}}) \quad (1 - t_{\tilde{p}}) \quad (1 - t_{\tilde{p}})^4 \\
r_2: & (1 - t_{\tilde{p}}^2)^4 = & (1 - t_{\tilde{p}}) \quad (1 - t_{\tilde{p}}) \quad (1 - t_{\tilde{p}}) \quad (1 - t_{\tilde{p}}) \quad (1 + t_{\tilde{p}})^4 \\
r, r_3: & (1 - t_{\tilde{p}}^4)^2 = & (1 - t_{\tilde{p}}) \quad (1 - t_{\tilde{p}}) \quad (1 + t_{\tilde{p}}) \quad (1 + t_{\tilde{p}}) \quad (1 + t_{\tilde{p}}^2)^2 \\
s, s_2: & (1 - t_{\tilde{p}}^2)^4 = & (1 - t_{\tilde{p}}) \quad (1 + t_{\tilde{p}}) \quad (1 - t_{\tilde{p}}) \quad (1 + t_{\tilde{p}}) \quad (1 - t_{\tilde{p}}^2)^2 \\
s_1, s_3: & (1 - t_{\tilde{p}}^2)^4 = & (1 - t_{\tilde{p}}) \quad (1 + t_{\tilde{p}}) \quad (1 + t_{\tilde{p}}) \quad (1 - t_{\tilde{p}}) \quad (1 - t_{\tilde{p}}^2)^2
\end{array}$$

The possible irreducible segment group elements $\mathbf{h}_{\tilde{p}}$ are listed in the first column; s, s_2 denotes a reflection across either the x-axis or the y-axis, and s_1, s_3 denotes a reflection across a diagonal (see figure A25.1). In addition, degenerate pairs of boundary orbits can run along the symmetry lines in the full space, with the fundamental domain group theory weights $\mathbf{h}_p = (r_2 + s)/2$ (axes) and $\mathbf{h}_p = (r_2 + s)/2$ (diagonals) respectively:

$$\begin{array}{lcl}
& A_1 & A_2 \quad B_1 \quad B_2 \quad E \quad (\text{A25.1}) \\
\text{axes:} & (1 - t_{\tilde{p}}^2)^2 = & (1 - t_{\tilde{p}})(1 - 0t_{\tilde{p}})(1 - t_{\tilde{p}})(1 - 0t_{\tilde{p}})(1 + t_{\tilde{p}})^2 \\
\text{diagonals:} & (1 - t_{\tilde{p}}^2)^2 = & (1 - t_{\tilde{p}})(1 - 0t_{\tilde{p}})(1 - 0t_{\tilde{p}})(1 - t_{\tilde{p}})(1 + t_{\tilde{p}})^2
\end{array}$$

(we have assumed that $t_{\tilde{p}}$ does not change sign under reflections across symmetry axes). For the 4-disk arrangement considered here only the diagonal orbits $\bar{13}, \bar{24}$ occur; they correspond to the $\bar{2}$ fixed point in the fundamental domain.

The A_1 subspace in D_4 cycle expansion is given by

$$\begin{aligned}
 1/\zeta_{A_1} &= (1-t_0)(1-t_1)(1-t_2)(1-t_{01})(1-t_{02})(1-t_{12}) \\
 &\quad (1-t_{001})(1-t_{002})(1-t_{011})(1-t_{012})(1-t_{021})(1-t_{022})(1-t_{112}) \\
 &\quad (1-t_{122})(1-t_{0001})(1-t_{0002})(1-t_{0011})(1-t_{0012})(1-t_{0021}) \dots \\
 &= 1 - t_0 - t_1 - t_2 - (t_{01} - t_0 t_1) - (t_{02} - t_0 t_2) - (t_{12} - t_1 t_2) \\
 &\quad - (t_{001} - t_0 t_{01}) - (t_{002} - t_0 t_{02}) - (t_{011} - t_1 t_{01}) \\
 &\quad - (t_{022} - t_2 t_{02}) - (t_{112} - t_1 t_{12}) - (t_{122} - t_2 t_{12}) \\
 &\quad - (t_{012} + t_{021} + t_0 t_1 t_2 - t_0 t_{12} - t_1 t_{02} - t_2 t_{01}) \dots
 \end{aligned} \tag{A25.2}$$

(for typographical convenience, $\underline{1}$ is replaced by 0 in the remainder of this section). For 1-dimensional representations, the characters can be read off the symbol strings: $\chi_{A_2}(\mathbf{h}_{\tilde{\mathbf{p}}}) = (-1)^{n_0}$, $\chi_{B_1}(\mathbf{h}_{\tilde{\mathbf{p}}}) = (-1)^{n_1}$, $\chi_{B_2}(\mathbf{h}_{\tilde{\mathbf{p}}}) = (-1)^{n_0+n_1}$, where n_0 and n_1 are the number of times symbols 0, 1 appear in the $\tilde{\mathbf{p}}$ symbol string. For B_2 all t_p with an odd total number of 0's and 1's change sign:

$$\begin{aligned}
 1/\zeta_{B_2} &= (1+t_0)(1+t_1)(1-t_2)(1-t_{01})(1+t_{02})(1+t_{12}) \\
 &\quad (1+t_{001})(1-t_{002})(1+t_{011})(1-t_{012})(1-t_{021})(1+t_{022})(1-t_{112}) \\
 &\quad (1+t_{122})(1-t_{0001})(1+t_{0002})(1-t_{0011})(1+t_{0012})(1+t_{0021}) \dots \\
 &= 1 + t_0 + t_1 - t_2 - (t_{01} - t_0 t_1) + (t_{02} - t_0 t_2) + (t_{12} - t_1 t_2) \\
 &\quad + (t_{001} - t_0 t_{01}) - (t_{002} - t_0 t_{02}) + (t_{011} - t_1 t_{01}) \\
 &\quad + (t_{022} - t_2 t_{02}) - (t_{112} - t_1 t_{12}) + (t_{122} - t_2 t_{12}) \\
 &\quad - (t_{012} + t_{021} + t_0 t_1 t_2 - t_0 t_{12} - t_1 t_{02} - t_2 t_{01}) \dots
 \end{aligned} \tag{A25.3}$$

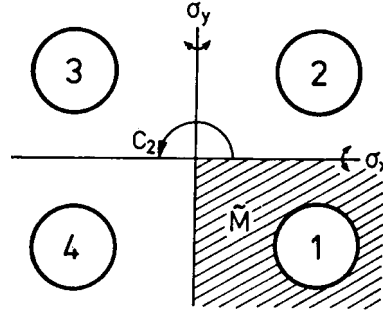
The form of the remaining cycle expansions depends crucially on the special role played by the boundary orbits: by (A25.1) the orbit t_2 does not contribute to A_2 and B_1 ,

$$\begin{aligned}
 1/\zeta_{A_2} &= (1+t_0)(1-t_1)(1+t_{01})(1+t_{02})(1-t_{12}) \\
 &\quad (1-t_{001})(1-t_{002})(1+t_{011})(1+t_{012})(1+t_{021})(1+t_{022})(1-t_{112}) \\
 &\quad (1-t_{122})(1+t_{0001})(1+t_{0002})(1-t_{0011})(1-t_{0012})(1-t_{0021}) \dots \\
 &= 1 + t_0 - t_1 + (t_{01} - t_0 t_1) + t_{02} - t_{12} \\
 &\quad - (t_{001} - t_0 t_{01}) - (t_{002} - t_0 t_{02}) + (t_{011} - t_1 t_{01}) \\
 &\quad + t_{022} - t_{122} - (t_{112} - t_1 t_{12}) + (t_{012} + t_{021} - t_0 t_{12} - t_1 t_{02}) .
 \end{aligned} \tag{A25.4}$$

and

$$\begin{aligned}
 1/\zeta_{B_1} &= (1-t_0)(1+t_1)(1+t_{01})(1-t_{02})(1+t_{12}) \\
 &\quad (1+t_{001})(1-t_{002})(1-t_{011})(1+t_{012})(1+t_{021})(1-t_{022})(1-t_{112}) \\
 &\quad (1+t_{122})(1+t_{0001})(1-t_{0002})(1-t_{0011})(1+t_{0012})(1+t_{0021}) \dots \\
 &= 1 - t_0 + t_1 + (t_{01} - t_0 t_1) - t_{02} + t_{12} \\
 &\quad + (t_{001} - t_0 t_{01}) - (t_{002} - t_0 t_{02}) - (t_{011} - t_1 t_{01}) \\
 &\quad - t_{022} + t_{122} - (t_{112} - t_1 t_{12}) + (t_{012} + t_{021} - t_0 t_{12} - t_1 t_{02}) .
 \end{aligned} \tag{A25.5}$$

Figure A25.3: Symmetries of four disks on a rectangle. A fundamental domain indicated by the shaded area.



In the above we have assumed that t_2 does not change sign under D_4 reflections. For the mixed-symmetry subspace E the curvature expansion is given by

$$\begin{aligned}
 1/\zeta_E = & 1 + t_2 + (-t_0^2 + t_1^2) + (2t_{002} - t_2t_0^2 - 2t_{112} + t_2t_1^2) \\
 & + (2t_{0011} - 2t_{0022} + 2t_2t_{002} - t_{01}^2 - t_{02}^2 + 2t_{1122} - 2t_2t_{112} \\
 & + t_{12}^2 - t_0^2t_1^2) + (2t_{00002} - 2t_{00112} + 2t_2t_{0011} - 2t_{00121} - 2t_{00211} \\
 & + 2t_{00222} - 2t_2t_{0022} + 2t_{01012} + 2t_{01021} - 2t_{01102} - t_2t_{01}^2 + 2t_{02022} \\
 & - t_2t_{02}^2 + 2t_{11112} - 2t_{11222} + 2t_2t_{1122} - 2t_{12122} + t_2t_{12}^2 - t_2t_0^2t_1^2 \\
 & + 2t_{002}(-t_0^2 + t_1^2) - 2t_{112}(-t_0^2 + t_1^2)) \quad (\text{A25.6})
 \end{aligned}$$

A quick test of the $\zeta = \zeta_{A_1}\zeta_{A_2}\zeta_{B_1}\zeta_{B_2}\zeta_E^2$ factorization is afforded by the topological polynomial; substituting $t_p = z^{n_p}$ into the expansion yields

$$1/\zeta_{A_1} = 1 - 3z, \quad 1/\zeta_{A_2} = 1/\zeta_{B_1} = 1, \quad 1/\zeta_{B_2} = 1/\zeta_E = 1 + z,$$

in agreement with (18.56).

exercise 23.8

A25.2 D_2 factorization

An arrangement of four identical disks on the vertices of a rectangle has D_2 symmetry, see figure A25.3. D_2 consists of $\{e, r, \sigma_x, \sigma_y\}$, i.e., a rotation by π and the reflections across the two symmetry axes.

This system affords a rather easy visualization of the conversion of a 4-disk dynamics into a fundamental domain symbolic dynamics. An orbit leaving the fundamental domain through one of the axis may be folded back by a reflection on that axis; with these symmetry operations $g_0 = s_x$ and $g_1 = s_y$ we associate labels 1 and 0, respectively. Orbits going to the diagonally opposed disk cross the boundaries of the fundamental domain twice; the product of these two reflections is just $r_2 = s_x s_y$, to which we assign the label 2. For example, a ternary string 0010201... is converted into 12143123..., and the associated group-theory weight is given by ... $g_1 g_0 g_2 g_0 g_1 g_0 g_0$.

Short ternary cycles and the corresponding 4-disk cycles are listed in table A25.2. Note that already at length three there is a pair of cycles (012 = 143 and 021 = 142) related by time reversal, but *not* by any D_2 symmetries.

Table A25.2: D_2 correspondence between the ternary $\{0, 1, 2\}$ fundamental domain prime cycles $\tilde{p}$ and the full 4-disk $\{1, 2, 3, 4\}$ cycles p , together with the D_2 transformation that maps the end point of the $\tilde{p}$ cycle into an irreducible segment of the p cycle. The degeneracy of the p cycle is $m_p = 4n_{\tilde{p}}/n_p$. Note that the 012 and 021 cycles are related by time reversal, but cannot be mapped into each other by D_2 transformations. The full space orbit listed here is generated from the symmetry reduced code by the rules given in sect. A25.2, starting from disk 1.

$\tilde{p}$	p	$\mathbf{g}$	$\tilde{p}$	p	$\mathbf{g}$
0	14	σ_y	0001	1414 3232	C_2
1	12	σ_x	0002	1414 2323	σ_x
2	13	r	0011	1412	e
01	14 32	r	0012	1412 4143	σ_y
02	14 23	σ_x	0021	1413 4142	σ_y
12	12 43	σ_y	0022	1413	e
001	141 232	σ_x	0102	1432 4123	σ_y
002	141 323	r	0111	1434 3212	r
011	143 412	σ_y	0112	1434 2343	σ_x
012	143	e	0121	1431 2342	σ_x
021	142	e	0122	1431 3213	r
022	142 413	σ_y	0211	1421 2312	σ_x
112	121 343	r	0212	1421 3243	r
122	124 213	σ_x	0221	1424 3242	r
			0222	1424 2313	σ_x
			1112	1212 4343	σ_y
			1122	1213	e
			1222	1242 4313	σ_y

The above is the complete description of the symbolic dynamics for 4 sufficiently separated equal disks placed at corners of a rectangle. However, if the fundamental domain requires further partitioning, the ternary description is insufficient. For example, in the stadium billiard fundamental domain one has to distinguish between bounces off the straight and the curved sections of the billiard wall; in that case five symbols suffice for constructing the covering symbolic dynamics.

The group D_2 has four 1-dimensional representations, distinguished by their behavior under axis reflections. The A_1 representation is symmetric with respect to both reflections; the A_2 representation is antisymmetric with respect to both. The B_1 and B_2 representations are symmetric under one and antisymmetric under the other reflection. The character table is

D_2	A_1	A_2	B_1	B_2
e	1	1	1	1
r_2	1	1	-1	-1
s_x	1	-1	1	-1
s_y	1	-1	-1	1

Substituted into the factorized determinant (25.19), the contributions of periodic orbits split as follows

$g_{\bar{p}}$		A_1	A_2	B_1	B_2
e :	$(1 - t_{\bar{p}})^4$	$=$	$(1 - t_{\bar{p}})$	$(1 - t_{\bar{p}})$	$(1 - t_{\bar{p}})$
r_2 :	$(1 - t_{\bar{p}}^2)^2$	$=$	$(1 - t_{\bar{p}})$	$(1 - t_{\bar{p}})$	$(1 - t_{\bar{p}})$
s_x :	$(1 - t_{\bar{p}}^2)^2$	$=$	$(1 - t_{\bar{p}})$	$(1 + t_{\bar{p}})$	$(1 + t_{\bar{p}})$
s_y :	$(1 - t_{\bar{p}}^2)^2$	$=$	$(1 - t_{\bar{p}})$	$(1 + t_{\bar{p}})$	$(1 - t_{\bar{p}})$

Cycle expansions follow by substituting cycles and their group theory factors from table A25.2. For A_1 all characters are +1, and the corresponding cycle expansion is given in (A25.2). Similarly, the totally antisymmetric subspace factorization A_2 is given by (A25.3), the B_2 factorization of D_4 . For B_1 all t_p with an odd total number of 0's and 2's change sign:

$$\begin{aligned}
 1/\zeta_{B_1} &= (1 + t_0)(1 - t_1)(1 + t_2)(1 + t_{01})(1 - t_{02})(1 + t_{12}) \\
 &\quad (1 - t_{001})(1 + t_{002})(1 + t_{011})(1 - t_{012})(1 - t_{021})(1 + t_{022})(1 + t_{112}) \\
 &\quad (1 - t_{122})(1 + t_{0001})(1 - t_{0002})(1 - t_{0011})(1 + t_{0012})(1 + t_{0021}) \dots \\
 &= 1 + t_0 - t_1 + t_2 + (t_{01} - t_0 t_1) - (t_{02} - t_0 t_2) + (t_{12} - t_1 t_2) \\
 &\quad -(t_{001} - t_0 t_{01}) + (t_{002} - t_0 t_{02}) + (t_{011} - t_1 t_{01}) \\
 &\quad + (t_{022} - t_2 t_{02}) + (t_{112} - t_1 t_{12}) - (t_{122} - t_2 t_{12}) \\
 &\quad -(t_{012} + t_{021} + t_0 t_1 t_2 - t_0 t_{12} - t_1 t_{02} - t_2 t_{01}) \dots
 \end{aligned} \tag{A25.7}$$

For B_2 all t_p with an odd total number of 1's and 2's change sign:

$$\begin{aligned}
 1/\zeta_{B_2} &= (1 - t_0)(1 + t_1)(1 + t_2)(1 + t_{01})(1 + t_{02})(1 - t_{12}) \\
 &\quad (1 + t_{001})(1 + t_{002})(1 - t_{011})(1 - t_{012})(1 - t_{021})(1 - t_{022})(1 + t_{112}) \\
 &\quad (1 + t_{122})(1 + t_{0001})(1 + t_{0002})(1 - t_{0011})(1 - t_{0012})(1 - t_{0021}) \dots \\
 &= 1 - t_0 + t_1 + t_2 + (t_{01} - t_0 t_1) + (t_{02} - t_0 t_2) - (t_{12} - t_1 t_2) \\
 &\quad + (t_{001} - t_0 t_{01}) + (t_{002} - t_0 t_{02}) - (t_{011} - t_1 t_{01}) \\
 &\quad - (t_{022} - t_2 t_{02}) + (t_{112} - t_1 t_{12}) + (t_{122} - t_2 t_{12}) \\
 &\quad -(t_{012} + t_{021} + t_0 t_1 t_2 - t_0 t_{12} - t_1 t_{02} - t_2 t_{01}) \dots
 \end{aligned} \tag{A25.8}$$

Note that all of the above cycle expansions group long orbits together with their pseudo-orbit shadows, so that the shadowing arguments for convergence still apply.

The topological polynomial factorizes as

$$\frac{1}{\zeta_{A_1}} = 1 - 3z, \quad \frac{1}{\zeta_{A_2}} = \frac{1}{\zeta_{B_1}} = \frac{1}{\zeta_{B_2}} = 1 + z,$$

consistent with the 4-disk factorization (18.56).

Commentary

Remark A25.1. D_4 labeling conventions While there is a variety of labeling conventions [2, 3] for the reduced D_4 dynamics, we prefer the one introduced here because

of its close relation to the group-theoretic structure of the dynamics: the global 4-disk trajectory can be generated by mapping the fundamental domain trajectories onto the full 4-disk space by the accumulated product of the D_4 group elements.

Remark A25.2. D_2 symmetry D_2 is the symmetry of several systems studied in the literature, such as the stadium billiard [1], and the 2-dimensional anisotropic Kepler potential [4].

References

- [1] L. A. Bunimovich, “On the ergodic properties of nowhere dispersing billiards”, *Commun. Math. Phys.* **65**, 295–312 (1979).
- [2] F. Christiansen, Analysis of Chaotic Dynamical Systems in Terms of Cycles, MA thesis (Univ. of Copenhagen, Copenhagen, 1989).
- [3] B. Eckhardt and D. Wintgen, “Symbolic description of periodic orbits for the quadratic Zeeman effect”, *J. Phys. B* **23**, 355–363 (1990).
- [4] M. C. Gutzwiller, “The quantization of a classically ergodic system”, *Physica D* **5**, 183–207 (1982).

Table A25.3: Character table of cyclic group C_n . Here $k, j = 1, 2, \dots, n-1$ and $\omega = e^{i2\pi/n}$ is an n th root of unity.

C_n	A	Γ_j
e	1	1
r_k	1	ω^{kj}

Table A25.4: Character table of dihedral group D_n , n odd. Here $k = 1, 2, \dots, (n-1)/2$.

D_n (n odd)	A_1	A_2	E_j
e	1	1	2
r_k, r_{n-k}	1	1	$\omega^{kj} + \omega^{-kj}$
$s, sr, \dots, sr_{n-1}$	1	-1	0

A25.3 Examples

Example A25.1. Character table of cyclic group C_n : The symmetry under a discrete rotation r by angle $2\pi/n$ gives birth to a cyclic group $C_n = \{e, r, r_2, \dots, r_{n-1}\}$. Since C_n is Abelian, each element forms a separate class, and thus C_n has n one-dimensional irreducible representations, with the character table [A25.3](#). The characters multiply as group elements: $\chi_\alpha(r_i)\chi_\alpha(r_j) = \chi_\alpha(r_{i+j})$. (Continued in example [A25.2](#))

(X. Ding)

Example A25.2. Character table of dihedral group D_n , n odd: (Continued from example [A25.1](#)) The D_n group

$$D_n = \{e, r, r_2, \dots, r_{n-1}, s, sr, \dots, sr_{n-1}\}$$

has n rotation elements and n reflections. Group elements satisfy $r_i r_j s = r_j sr_{n-i}$, so a clockwise rotation r_i and the corresponding anticlockwise rotation r_{n-i} form a class. As $r_{n-i} r_{2i+j} s = r_j sr_{n-i}$, $r_j s$ and $r_{2i+j} s$ are in the same class. Therefore, there are only three different types of classes: $\{e\}$, $\{r_k, r_{n-k}\}$ and $\{s, sr, \dots, sr_{n-1}\}$. The total number of classes is $(n+3)/2$. So for n odd, there are 2 one-dimensional irreducible representations (symmetric A_1 and anti-symmetric A_2) and $(n-1)/2$ two-dimensional irreducible representations, with the character table [A25.4](#).

In the j th two-dimensional irreducible representation, the 3 kinds of classes have form

class/irrep (n odd)	E_j
e	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
r_k, r_{n-k}	$\begin{pmatrix} \omega^{kj} & 0 \\ 0 & \omega^{-kj} \end{pmatrix}$
$s, sr, \dots, sr_{n-1}$	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

(Continued in example [A25.3](#))

(X. Ding)

Example A25.3. Character table of dihedral group D_n , n even: (Continued from example [A25.2](#)) In this case, there are $(n+6)/2$ classes: $\{e\}$, $\{r_{n/2}\}$, $\{r_k, r_{n-k}\}$, $\{s, sr_2, \dots, sr_{n-2}\}$

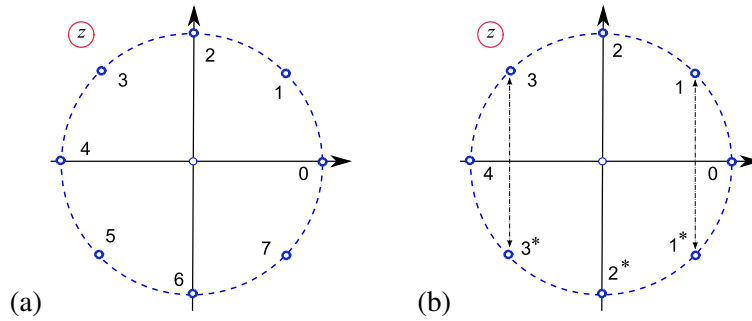


Figure A25.4: (a) As irreps of any C_n are one-dimensional, the complex roots of unity are a faithful representation of C_8 , and every root is a class by itself. (b) In the complex C_8 plane, the discrete symmetry of D_4 is given by complex conjugation, with 2-element classes $\{r_k, r_{n-k}\}$.

Table A25.5: Character table of dihedral group D_n , n even. Here $k, j = 1, 2, \dots, (n-2)/2$. See also figure A25.4.

D_n (n even)	A_1	A_2	B_1	B_2	E_j
e	1	1	1	1	2
$r_{n/2}$	1	1	$(-1)^{n/2}$	$(-1)^{n/2}$	$2(-1)^j$
r_k, r_{n-k} (k odd)	1	1	-1	-1	$\omega^{kj} + \omega^{-kj}$
r_k, r_{n-k} (k even)	1	1	1	1	$\omega^{kj} + \omega^{-kj}$
$s, sr_2, \dots, sr_{n-2}$	1	-1	1	-1	0
$sr, sr_3, \dots, sr_{n-1}$	1	-1	-1	1	0

and $\{sr, sr_3, \dots, sr_{n-1}\}$. There are four different one-dimensional irreducible representations, whose characters are ± 1 under reflection s and shift-reflect operation sr . We get table A25.5. (X. Ding)

Appendix A31

Koopman modes

A31.1 Koopmania

THE KOOPMAN OPERATOR action (31.1) on a state space function $a(x)$ is to replace it by its downstream value time t later, $a(x) \rightarrow a(x(t))$, evaluated at the trajectory point $x(t)$:

$$\begin{aligned} [\mathcal{K}^t a](x) &= a(f^t(x)) = \int_{\mathcal{M}} dy \mathcal{K}^t(x, y) a(y) \\ \mathcal{K}^t(x, y) &= \delta(y - f^t(x)) . \end{aligned} \tag{A31.1}$$

Eq. (31.2) suggests an alternative point of view, which is to push dynamical effects into the density. In contrast to the Koopman operator which advances the trajectory by time t , the Perron-Frobenius operator depends on the trajectory point time t in the past

Here we limit ourselves to a brief remark about the notion of “spectrum” of a linear operator.

The Koopman operator $\mathcal{K}$ acts multiplicatively in time, so it is reasonable to suppose that there exist constants $M > 0$, $\beta \geq 0$ such that $\|\mathcal{K}^t\| \leq M e^{t\beta}$ for all $t \geq 0$. What does that mean? The operator norm is define in the same spirit in which we defined the matrix norms in sect. A40.2: We are assuming that no value of $\mathcal{K}^t \rho(x)$ grows faster than exponentially for any choice of function $\rho(x)$, so that the fastest possible growth can be bounded by $e^{t\beta}$, a reasonable expectation in the light of the simplest example studied so far, the exact escape rate (20.31). If that is so, multiplying $\mathcal{K}^t$ by $e^{-t\beta}$ we construct a new operator $e^{-t\beta} \mathcal{K}^t = e^{t(\mathcal{A}-\beta)}$ which decays exponentially for large t , $\|e^{t(\mathcal{A}-\beta)}\| \leq M$. We say that $e^{-t\beta} \mathcal{K}^t$ is an element of a *bounded* semigroup with generator $\mathcal{A} - \beta \mathbf{1}$. Given this bound, it follows by

the Laplace transform

$$\int_0^\infty dt e^{-st} \mathcal{K}^t = \frac{1}{s - \mathcal{A}}, \quad \operatorname{Re} s > \beta, \quad (\text{A31.2})$$

that the *resolvent* operator $(s - \mathcal{A})^{-1}$ is bounded (“resolvent” = able to cause separation into constituents)

section A40.2

remark A22.1

$$\left\| \frac{1}{s - \mathcal{A}} \right\| \leq \int_0^\infty dt e^{-st} M e^{t\beta} = \frac{M}{s - \beta}.$$

If one is interested in the spectrum of $\mathcal{K}$, as we will be, the resolvent operator is a natural object to study. The main lesson of this brief aside is that for the continuous time flows the Laplace transform is the tool that brings down the generator in (19.27) into the resolvent form (20.22) and enables us to study its spectrum.

A31.2 Implementing evolution

(R. Artuso and P. Cvitanović)



We now come back to the semigroup of operators $\mathcal{K}^t$. We have introduced the generator of the semigroup (19.25) as

$$\mathcal{A} = \left. \frac{d}{dt} \mathcal{K}^t \right|_{t=0}.$$

If we now take the derivative at arbitrary times we get

$$\begin{aligned} \left(\frac{d}{dt} \mathcal{K}^t \psi \right)(x) &= \lim_{\eta \rightarrow 0} \frac{\psi(f^{t+\eta}(x)) - \psi(f^t(x))}{\eta} \\ &= v_i(f^t(x)) \left. \frac{\partial}{\partial \tilde{x}_i} \psi(\tilde{x}) \right|_{\tilde{x}=f^t(x)} \\ &= (\mathcal{K}^t \mathcal{A} \psi)(x) \end{aligned}$$

which can be formally integrated like an ordinary differential equation yielding

exercise A31.1

$$\mathcal{K}^t = e^{t\mathcal{A}}. \quad (\text{A31.3})$$

This guarantees that the Laplace transform manipulations in sect. 19.5 are correct. Though the formal expression of the semigroup (A31.3) is quite simple one has to take care in implementing its action. If we express the exponential through the power series

$$\mathcal{K}^t = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathcal{A}^k, \quad (\text{A31.4})$$

we encounter the problem that the infinitesimal generator (19.25) contains non-commuting pieces, i.e., there are i, j combinations for which the commutator does not satisfy

$$\left[\frac{\partial}{\partial x_i}, v_j(x) \right] = 0.$$

To derive a more useful representation, we follow the strategy used for finite-dimensional matrix operators in sects. 4.3 and 4.4 and use the semigroup property to write

$$\mathcal{K}^t = \prod_{m=1}^{t/\delta\tau} \mathcal{K}^{\delta\tau}$$

as the starting point for a discretized approximation to the continuous time dynamics, with time step $\delta\tau$. Omitting terms from the second order onwards in the expansion of $\mathcal{K}^{\delta\tau}$ yields an error of order $O(\delta\tau^2)$. This might be acceptable if the time step $\delta\tau$ is sufficiently small. In practice we write the Euler product

$$\mathcal{K}^t = \prod_{m=1}^{t/\delta\tau} (1 + \delta\tau \mathcal{A}_{(m)}) + O(\delta\tau^2) \quad (\text{A31.5})$$

where

$$(\mathcal{A}_{(m)}\psi)(x) = v_i(f^{m\delta\tau}(x)) \left. \frac{\partial\psi}{\partial x_i} \right|_{\tilde{x}=f^{m\delta\tau}(x)}$$

As far as the x dependence is concerned, $e^{\delta\tau \mathcal{A}_i}$ acts as

$$e^{\delta\tau \mathcal{A}_i} \begin{Bmatrix} x_1 \\ \cdot \\ x_i \\ \cdot \\ x_d \end{Bmatrix} \rightarrow \begin{Bmatrix} x_1 \\ \cdot \\ x_i + \delta\tau v_i(x) \\ \cdot \\ x_d \end{Bmatrix}. \quad (\text{A31.6})$$

We see that the product form (A31.5) of the operator is nothing else but a prescription for finite time step integration of the equations of motion - in this case the simplest Euler type integrator which advances the trajectory by $\delta\tau \times \text{velocity}$ at each time step.

exercise 2.6

A31.2.1 A symplectic integrator



The procedure we described above is only a starting point for more sophisticated approximations. As an example on how to get a sharper bound on the error term consider the Hamiltonian flow $\mathcal{A} = \mathcal{B} + \mathcal{C}$, $\mathcal{B} = p_i \frac{\partial}{\partial q_i}$, $\mathcal{C} = -\partial_i V(q) \frac{\partial}{\partial p_i}$. Clearly the potential and the kinetic parts do not commute. We make sense of the formal solution (A31.5) by splitting it into infinitesimal steps and keeping terms up to $\delta\tau^2$ in

exercise A31.2

$$\mathcal{K}^{\delta\tau} = \hat{\mathcal{K}}^{\delta\tau} + \frac{1}{24} \delta\tau^3 [\mathcal{B} + 2\mathcal{C}, [\mathcal{B}, \mathcal{C}]] + \cdots, \quad (\text{A31.7})$$

where

$$\hat{\mathcal{K}}^{\delta\tau} = e^{\frac{1}{2}\delta\tau\mathcal{B}} e^{\delta\tau\mathcal{C}} e^{\frac{1}{2}\delta\tau\mathcal{B}}. \quad (\text{A31.8})$$

The approximate infinitesimal Liouville operator $\hat{\mathcal{K}}^{\delta\tau}$ is of the form that now generates evolution as a sequence of mappings induced by (19.28), a free flight by $\frac{1}{2}\delta\tau\mathcal{B}$, scattering by $\delta\tau\partial V(q')$, followed again by $\frac{1}{2}\delta\tau\mathcal{B}$ free flight:

$$\begin{aligned} e^{\frac{1}{2}\delta\tau\mathcal{B}} \begin{Bmatrix} q \\ p \end{Bmatrix} &\rightarrow \begin{Bmatrix} q' \\ p' \end{Bmatrix} = \begin{Bmatrix} q - \frac{\delta\tau}{2}p \\ p \end{Bmatrix} \\ e^{\delta\tau\mathcal{C}} \begin{Bmatrix} q' \\ p' \end{Bmatrix} &\rightarrow \begin{Bmatrix} q'' \\ p'' \end{Bmatrix} = \begin{Bmatrix} q' \\ p' + \delta\tau\partial V(q') \end{Bmatrix} \\ e^{\frac{1}{2}\delta\tau\mathcal{B}} \begin{Bmatrix} q'' \\ p'' \end{Bmatrix} &\rightarrow \begin{Bmatrix} q''' \\ p''' \end{Bmatrix} = \begin{Bmatrix} q' - \frac{\delta\tau}{2}p'' \\ p'' \end{Bmatrix} \end{aligned} \quad (\text{A31.9})$$

Collecting the terms we obtain an integration rule for this type of symplectic flow which is better than the straight Euler integration (A31.6) as it is accurate up to order $\delta\tau^2$:

$$\begin{aligned} q_{n+1} &= q_n - \delta\tau p_n - \frac{(\delta\tau)^2}{2} \partial V(q_n - \delta\tau p_n/2) \\ p_{n+1} &= p_n + \delta\tau \partial V(q_n - \delta\tau p_n/2) \end{aligned} \quad (\text{A31.10})$$

The Jacobian matrix of one integration step is given by

$$M = \begin{bmatrix} 1 & -\delta\tau/2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \delta\tau\partial V(q') & 1 \end{bmatrix} \begin{bmatrix} 1 & -\delta\tau/2 \\ 0 & 1 \end{bmatrix}. \quad (\text{A31.11})$$

Note that the billiard flow (9.10) is an example of such symplectic integrator. In that case the free flight is interrupted by instantaneous wall reflections, and can be integrated out.

Commentary

Remark A31.1. Symplectic integration. The reviews [1] and [2] offer a good starting point for exploring the symplectic integrators literature. For a higher order integrators of type (A31.8), check ref. [3].

References

- [1] P. J. Channell and C. Scovel, “Symplectic integration of Hamiltonian systems”, *Nonlinearity* **3**, 231–259 (1990).
- [2] J. M. Sanz-Serna and M. P. Calvo, *Numerical Hamiltonian problems* (Chapman and Hall, London UK, 1994).
- [3] M. Suzuki, “General theory of fractal path integrals with applications to many-body theories and statistical physics”, *J. Math. Phys.* **32**, 400 (1991).

Exercises

- A31.1. **Exponential form of semigroup elements.** Check that the Koopman operator and the evolution generator commute, $\mathcal{K}^t \mathcal{A} = \mathcal{A} \mathcal{K}^t$, by considering the action of both operators on an arbitrary state space function $a(x)$.
- A31.2. **Non-commutativity.** Check that the commutators in (A31.7) are not vanishing by showing that

$$[\mathcal{B}, \mathcal{C}] = -p \left(V'' \frac{\partial}{\partial p} - V' \frac{\partial}{\partial q} \right).$$

- A31.3. **Symplectic leapfrog integrator.** Implement (A31.10) for 2-dimensional Hamiltonian flows; compare it with Runge-Kutta integrator by integrating trajectories in some (chaotic) Hamiltonian flow.
- A31.4. **Symplectic volume preservation.** Check that the sequence of mappings (A31.9) is volume preserving, $\det \hat{U} = 1$.

Appendix A32

Thermodynamic formalism

Being Hungarian is not sufficient. You also must be talented.

— Zsa Zsa Gabor

(G. Vattay)

IN THE PRECEDING CHAPTERS we characterized chaotic systems via global quantities such as averages. It turned out that these are closely related to very fine details of the dynamics like stabilities and time periods of individual periodic orbits. In statistical mechanics a similar duality exists. Macroscopic systems are characterized with thermodynamic quantities (pressure, temperature and chemical potential) which are averages over fine details of the system called microstates. One of the greatest achievements of the theory of dynamical systems was when in the sixties and seventies Bowen, Ruelle and Sinai made the analogy between these two subjects explicit. Later this “Thermodynamic Formalism” of dynamical systems became widely used making it possible to calculate various fractal dimensions. We sketch the main ideas of this theory and show how periodic orbit theory helps to carry out calculations.

A32.1 Rényi entropies

As we have already seen trajectories in a dynamical system can be characterized by their symbolic sequences from a generating Markov partition. We can locate the set of starting points $\mathcal{M}_{s_1 s_2 \dots s_n}$ of trajectories whose symbol sequence starts with a given set of n symbols $s_1 s_2 \dots s_n$. We can associate many different quantities to these sets. There are geometric measures such as the volume $V(s_1 s_2 \dots s_n)$, the area $A(s_1 s_2 \dots s_n)$ or the length $l(s_1 s_2 \dots s_n)$ of this set. Or in general we can have some measure $\mu(\mathcal{M}_{s_1 s_2 \dots s_n}) = \mu(s_1 s_2 \dots s_n)$ of this set. As we have seen in (27.10) the most important is the natural measure, which is the probability that an ergodic trajectory visits the set $\mu(s_1 s_2 \dots s_n) = P(s_1 s_2 \dots s_n)$. The natural measure is additive.

Summed up for all possible symbol sequences of length n it gives the measure of the whole state space:

$$\sum_{s_1 s_2 \dots s_n} \mu(s_1 s_2 \dots s_n) = 1 \quad (\text{A32.1})$$

expresses probability conservation. Also, summing up for the last symbol we get the measure of a one step shorter sequence

$$\sum_{s_n} \mu(s_1 s_2 \dots s_n) = \mu(s_1 s_2 \dots s_{n-1}).$$

As we increase the length (n) of the sequence the measure associated with it decreases typically with an exponential rate. It is then useful to introduce the exponents

$$\lambda(s_1 s_2 \dots s_n) = -\frac{1}{n} \log \mu(s_1 s_2 \dots s_n). \quad (\text{A32.2})$$

To get full information on the distribution of the natural measure in the symbolic space we can study the distribution of exponents. Let the number of symbol sequences of length n with exponents between λ and $\lambda + d\lambda$ be given by $N_n(\lambda)d\lambda$. For large n the number of such sequences increases exponentially. The rate of this exponential growth can be characterized by $g(\lambda)$ such that

$$N_n(\lambda) \sim \exp(ng(\lambda)).$$

The knowledge of the distribution $N_n(\lambda)$ or its essential part $g(\lambda)$ fully characterizes the microscopic structure of our dynamical system.

As a natural next step we would like to calculate this distribution. However it is very time consuming to calculate the distribution directly by making statistics for millions of symbolic sequences. Instead, we introduce auxiliary quantities which are easier to calculate and to handle. These are called partition sums

$$Z_n(\beta) = \sum_{s_1 s_2 \dots s_n} \mu^\beta(s_1 s_2 \dots s_n), \quad (\text{A32.3})$$

as they are obviously motivated by Gibbs type partition sums of statistical mechanics. The parameter β plays the role of inverse temperature $1/k_B T$ and $E(s_1 s_2 \dots s_n) = -\log \mu(s_1 s_2 \dots s_n)$ is the energy associated with the microstate labeled by $s_1 s_2 \dots s_n$. We are tempted also to introduce something analogous with the Free energy. In dynamical systems this is called the Rényi entropy [1] defined by the growth rate of the partition sum

$$K_\beta = \lim_{n \rightarrow \infty} \frac{1}{n} \frac{1}{1 - \beta} \log \left(\sum_{s_1 s_2 \dots s_n} \mu^\beta(s_1 s_2 \dots s_n) \right). \quad (\text{A32.4})$$

In the special case $\beta \rightarrow 1$ we get Kolmogorov entropy

$$K_1 = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{s_1 s_2 \dots s_n} -\mu(s_1 s_2 \dots s_n) \log \mu(s_1 s_2 \dots s_n),$$

while for $\beta = 0$ we recover the topological entropy

$$h_{top} = K_0 = \lim_{n \rightarrow \infty} \frac{1}{n} \log N(n),$$

where $N(n)$ is the number of existing length n sequences. To connect the partition sums with the distribution of the exponents, we can write them as averages over the exponents

$$Z_n(\beta) = \int d\lambda N_n(\lambda) \exp(-n\lambda\beta),$$

where we used the definition (A32.2). For large n we can replace $N_n(\lambda)$ with its asymptotic form

$$Z_n(\beta) \sim \int d\lambda \exp(ng(\lambda)) \exp(-n\lambda\beta).$$

For large n this integral is dominated by contributions from those λ^* which maximize the exponent

$$g(\lambda) - \lambda\beta.$$

The exponent is maximal when the derivative of the exponent vanishes

$$g'(\lambda^*) = \beta. \quad (\text{A32.5})$$

From this equation we can determine $\lambda^*(\beta)$. Finally the partition sum is

$$Z_n(\beta) \sim \exp(n[g(\lambda^*(\beta)) - \lambda^*(\beta)\beta]).$$

Using the definition (A32.4) we can now connect the Rényi entropies and $g(\lambda)$

$$(\beta - 1)K_\beta = \lambda^*(\beta)\beta - g(\lambda^*(\beta)). \quad (\text{A32.6})$$

Equations (A32.5) and (A32.6) define the Legendre transform of $g(\lambda)$. This equation is analogous with the thermodynamic equation connecting the entropy and the free energy. As we know from thermodynamics we can invert the Legendre transform. In our case we can express $g(\lambda)$ from the Rényi entropies via the Legendre transformation

$$g(\lambda) = \lambda\beta^*(\lambda) - (\beta^*(\lambda) - 1)K_{\beta^*(\lambda)}, \quad (\text{A32.7})$$

where now $\beta^*(\lambda)$ can be determined from

$$\frac{d}{d\beta^*}[(\beta^* - 1)K_{\beta^*}] = \lambda. \quad (\text{A32.8})$$

Obviously, if we can determine the Rényi entropies we can recover the distribution of probabilities from (A32.7) and (A32.8).

The periodic orbit calculation of the Rényi entropies can be carried out by approximating the natural measure corresponding to a symbol sequence by the expression (27.10)

$$\mu(s_1, \dots, s_n) \approx \frac{e^{n\gamma}}{|\Lambda_{s_1 s_2 \dots s_n}|}. \quad (\text{A32.9})$$

The partition sum (A32.3) now reads

$$Z_n(\beta) \approx \sum_i \frac{e^{n\beta\gamma}}{|\Lambda_i|^\beta}, \quad (\text{A32.10})$$

where the summation goes for periodic orbits of period n . We can define the characteristic function

$$\Omega(z, \beta) = \exp \left(- \sum_n \frac{z^n}{n} Z_n(\beta) \right). \quad (\text{A32.11})$$

According to (A32.4) for large n the partition sum behaves as

$$Z_n(\beta) \sim e^{-n(\beta-1)K_\beta}. \quad (\text{A32.12})$$

Substituting this into (A32.11) we can see that the leading zero of the characteristic function is

$$z_0(\beta) = e^{(\beta-1)K_\beta}.$$

On the other hand substituting the periodic orbit approximation (A32.10) into (A32.11) and introducing prime and repeated periodic orbits as usual we get

$$\Omega(z, \beta) = \exp \left(- \sum_{p,r} \frac{z^{n_{pr}} e^{\beta\gamma n_{pr}}}{r |\Lambda_p^r|^\beta} \right).$$

We can see that the characteristic function is the same as the zeta function we introduced for Lyapunov exponents (A6.12) except we have $ze^{\beta\gamma}$ instead of z . Then we can conclude that the Rényi entropies can be expressed with the pressure function directly as

$$P(\beta) = (\beta - 1)K_\beta + \beta\gamma, \quad (\text{A32.13})$$

since the leading zero of the zeta function is the pressure. The Rényi entropies K_β , hence the distribution of the exponents $g(\lambda)$ as well, can be calculated via finding the leading eigenvalue of the operator (A6.4).

From (A32.13) we can get all the important quantities of the thermodynamic formalism. For $\beta = 0$ we get the topological entropy

$$P(0) = -K_0 = -h_{top}. \quad (\text{A32.14})$$

For $\beta = 1$ we get the escape rate

$$P(1) = \gamma. \quad (\text{A32.15})$$

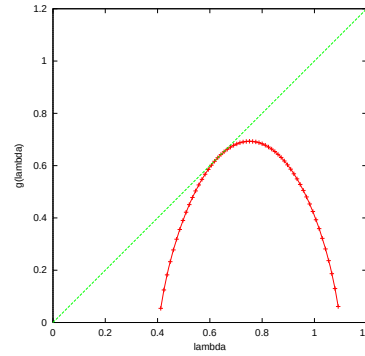
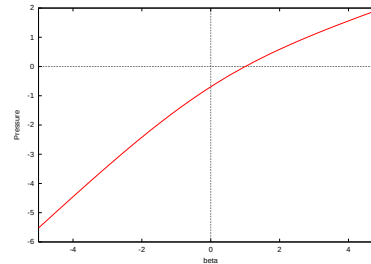
Taking the derivative of (A32.13) in $\beta = 1$ we get Pesin's formula [5] connecting Kolmogorov entropy and the Lyapunov exponent

$$P'(1) = \bar{\lambda} = K_1 + \gamma. \quad (\text{A32.16})$$

It is important to note that, as always, these formulas are strictly valid for nice

exercise A32.1

Figure A32.1

Figure A32.2: $g(\lambda)$ and $P(\beta)$ for the map of exercise A32.4 at $a = 3$ and $b = 3/2$.

hyperbolic systems only. At the end of this Chapter we discuss the important problems we are facing in nonhyperbolic cases.

In figure A32.2 we show a typical pressure and $g(\lambda)$ curve computed for the two scale tent map of exercise A32.4. We have to mention, that all typical hyperbolic dynamical system produces a similar parabola like curve. Although this is somewhat boring we can interpret it like a sign of a high level of universality: The exponents λ have a sharp distribution around the most probable value. The most probable value is $\lambda = P'(0)$ and $g(\lambda) = h_{top}$ is the topological entropy. The average value in closed systems is where $g(\lambda)$ touches the diagonal: $\bar{\lambda} = g(\bar{\lambda})$ and $1 = g'(\bar{\lambda})$.

Next, we are looking at the distribution of trajectories in real space.

A32.2 Fractal dimensions

Hentschel and Procaccia rediscovered a small part of my theory. Generalized dimensions are not useful at all.

—Benoit B. Mandelbrot

By looking at the repeller we can recognize an interesting spatial structure. In the 3-disk case the starting points of trajectories not leaving the system after the first bounce form two strips. Then these strips are subdivided into an infinite hierarchy of substrings as we follow trajectories which do not leave the system after more and more bounces. The finer strips are similar to strips on a larger scale. Objects with such self similar properties are called *fractals*.

We can characterize fractals via their local scaling properties. The first step is

to draw a uniform grid on the surface of section. We can look at various measures in the square boxes of the grid. The most interesting measure is again the natural measure located in the box. By decreasing the size of the grid ϵ the measure in a given box will decrease. If the distribution of the measure is smooth then we expect that the measure of the i th box is proportional with the dimension of the section

$$\mu_i \sim \epsilon^d.$$

If the measure is distributed on a hairy object like the repeller we can observe unusual scaling behavior of type

$$\mu_i \sim \epsilon^{\alpha_i},$$

where α_i is the local “dimension” or Hölder exponent of the object. As α is not necessarily an integer here we are dealing with objects with fractional dimensions. We can study the distribution of the measure on the surface of section by looking at the distribution of these local exponents. We can define

$$\alpha_i = \frac{\log \mu_i}{\log \epsilon},$$

the local Hölder exponent and then we can count how many of them are between α and $\alpha + d\alpha$. This is $N_\epsilon(\alpha)d\alpha$. Again, in smooth objects this function scales simply with the dimension of the system

$$N_\epsilon(\alpha) \sim \epsilon^{-d},$$

while for hairy objects we expect an α dependent scaling exponent

$$N_\epsilon(\alpha) \sim \epsilon^{-f(\alpha)}.$$

$f(\alpha)$ can be interpreted as the dimension of the points on the surface of section with scaling exponent α . [A note to the reader: Even though the thermodynamic formalism is of older vintage (we refer the reader to ref. [6] for a comprehensive overview), we adhere here to the notational conventions of ref. [4] which are more current in the physics literature: we strongly recommend also ref. [10], dealing with period doubling universality.] We can calculate $f(\alpha)$ with the help of partition sums as we did for $g(\lambda)$ in the previous section. First, we define

$$Z_\epsilon(q) = \sum_i \mu_i^q. \quad (\text{A32.17})$$

Then we would like to determine the asymptotic behavior of the partition sum characterized by the $\tau(q)$ exponent

$$Z_\epsilon(q) \sim \epsilon^{-\tau(q)}.$$

The partition sum can be written in terms of the distribution function of α -s

$$Z_\epsilon(q) = \int d\alpha N_\epsilon(\alpha) \epsilon^{q\alpha}.$$

Using the asymptotic form of the distribution we get

$$Z_\epsilon(q) \sim \int d\alpha \epsilon^{q\alpha - f(\alpha)}.$$

As ϵ goes to zero the integral is dominated by the term maximizing the exponent. This α^* can be determined from the equation

$$\frac{d}{d\alpha^*}(q\alpha^* - f(\alpha^*)) = 0,$$

leading to

$$q = f'(\alpha^*).$$

Finally we can read off the scaling exponent of the partition sum

$$\tau(q) = \alpha^* q - f(\alpha^*).$$

In a uniform fractal characterized by a single dimension both α and $f(\alpha)$ collapse to $\alpha = f(\alpha) = D$. The scaling exponent then has the form $\tau(q) = (q - 1)D$. In case of non uniform fractals we can introduce generalized dimensions [3] D_q via the definition

$$D_q = \tau(q)/(q - 1).$$

Some of these dimensions have special names. For $q = 0$ the partition sum (A32.17) counts the number of non empty boxes $\tilde{N}_\epsilon$. Consequently

$$D_0 = - \lim_{\epsilon \rightarrow 0} \frac{\log \tilde{N}_\epsilon}{\log \epsilon},$$

is called the box counting dimension. For $q = 1$ the dimension can be determined as the limit of the formulas for $q \rightarrow 1$ leading to

$$D_1 = \lim_{\epsilon \rightarrow 0} \sum_i \mu_i \log \mu_i / \log \epsilon.$$

This is the scaling exponent of the Shannon information entropy [9] of the distribution, hence its name is *information dimension*.

Using equisize grids is impractical in most of the applications. Instead, we can rewrite (A32.17) into the more convenient form

$$\sum_i \frac{\mu_i^q}{\epsilon^{\tau(q)}} \sim 1. \quad (\text{A32.18})$$

If we cover the i th branch of the fractal with a grid of size l_i instead of ϵ we can use the relation [4]

$$\sum_i \frac{\mu_i^q}{l_i^{\tau(q)}} \sim 1, \quad (\text{A32.19})$$

the non-uniform grid generalization of A32.18. Next we show how can we use the periodic orbit formalism to calculate fractal dimensions. We have already seen that the width of the strips of the repeller can be approximated with the stabilities of the periodic orbits placed within them

$$l_i \sim \frac{1}{|\Lambda_i|}.$$

Then using this relation and the periodic orbit expression of the natural measure we can write (A32.19) into the form

$$\sum_i \frac{e^{q\gamma n}}{|\Lambda_i|^{q-\tau(q)}} \sim 1, \quad (\text{A32.20})$$

where the summation goes for periodic orbits of period n . The sum for stabilities can be expressed with the pressure function again

$$\sum_i \frac{1}{|\Lambda_i|^{q-\tau(q)}} \sim e^{-nP(q-\tau(q))},$$

and (A32.20) can be written as

$$e^{q\gamma n} e^{-nP(q-\tau(q))} \sim 1,$$

for large n . Finally we get an implicit formula for the dimensions

$$P(q - (q - 1)D_q) = q\gamma. \quad (\text{A32.21})$$

Solving this equation directly gives us the partial dimensions of the fractal repeller along the stable direction. We can see again that the pressure function alone contains all the relevant information. Setting $q = 0$ in (A32.21) we can prove that the zero of the pressure function is the box-counting dimension of the repeller

$$P(D_0) = 0.$$

Taking the derivative of (A32.21) in $q = 1$ we get

$$P'(1)(1 - D_1) = \gamma.$$

This way we can express the information dimension with the escape rate and the Lyapunov exponent

$$D_1 = 1 - \gamma/\bar{\lambda}. \quad (\text{A32.22})$$

If the system is bound ($\gamma = 0$) the information dimension and all other dimensions are $D_q = 1$. Also since $D_1 0$ is positive (A32.22) proves that the Lyapunov exponent must be larger than the escape rate $\bar{\lambda} > \gamma$ in general.

exercise A32.4
exercise A32.5

Résumé

In this chapter we have shown that thermodynamic quantities and various fractal dimensions can be expressed in terms of the pressure function. The pressure

function is the leading eigenvalue of the operator which generates the Lyapunov exponent. In the Lyapunov case β is just an auxiliary variable. In thermodynamics it plays an essential role. The good news of the chapter is that the distribution of locally fluctuating exponents should not be computed via making statistics. We can use cyclist formulas for determining the pressure. Then the pressure can be found using short cycles + curvatures. Here the head reaches the tail of the snake. We just argued that the statistics of long trajectories coded in $g(\lambda)$ and $P(\beta)$ can be calculated from short cycles. To use this intimate relation between long and short trajectories effectively is still a research level problem.

Commentary

Remark A32.1. ‘Thermodynamic formalism’. David Ruelle introduced the term ‘thermodynamic formalism’ [7, 8] in his 1978 book [6] *Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics*.

Remark A32.2. Mild phase transition. In nonhyperbolic systems the formulas derived in this chapter should be modified. As we mentioned in remark 27.1 in nonhyperbolic systems the periodic orbit expression of the measure can be

$$\mu_0 = e^{\gamma n} / |\Lambda_0|^\delta,$$

where δ can differ from 1. Usually it is $1/2$. For sufficiently *negative* β the corresponding term $1/|\Lambda_0|^\beta$ can dominate (A32.10) while in (A32.3) $e^{\gamma n} / |\Lambda_0|^{\delta\beta}$ plays no dominant role. In this case the pressure as a function of β can have a kink at the critical point $\beta = \beta_c$ where $\beta_c \log |\Lambda_0| = (\beta_c - 1)K_{\beta_c} + \beta_c \gamma$. For $\beta < \beta_c$ the pressure and the Rényi entropies differ

$$P(\beta) \neq (\beta - 1)K_\beta + \beta\gamma.$$

This phenomena is called phase transition. This is however not a very deep problem. We can fix the relation between pressure and the entropies by replacing $1/|\Lambda_0|$ with $1/|\Lambda_0|^\delta$ in (A32.10).

Remark A32.3. Hard phase transition. The really deep trouble of thermodynamics is caused by intermittency. In that case we have periodic orbits with $|\Lambda_0| \rightarrow 1$ as $n \rightarrow \infty$. Then for $\beta > 1$ the contribution of these orbits dominate both (A32.10) and (A32.3). Consequently the partition sum scales as $Z_n(\beta) \rightarrow 1$ and both the pressure and the entropies are zero. In this case quantities connected with $\beta \leq 1$ make sense only. These are for example the topological entropy, Kolmogorov entropy, Lyapunov exponent, escape rate, D_0 and D_1 . This phase transition cannot be fixed. It is probably fair to say that quantities which depend on this phase transition are only of mathematical interest and not very useful for characterization of realistic dynamical systems.

References

- [1] J. Balatoni and A. Rényi, “Remarks on entropy”, Publ. Math. Inst. Hung. Acad. Sci. **1**, 9–40 (1956).

- [2] O. Biham and M. Kvale, “Unstable periodic orbits in the stadium billiard”, *Phys. Rev. A* **46**, 6334–6339 (1992).
- [3] P. Grassberger, “Generalized dimensions of strange attractors”, *Phys. Lett. A* **97**, 227–230 (1983).
- [4] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman, “Fractal measures and their singularities: The characterization of strange sets”, *Phys. Rev. A* **33**, 1141–1151 (1986).
- [5] Y. B. Pesin, “Characteristic Lyapunov exponents and smooth ergodic theory”, *Russian Math. Surveys* **32**, 55–114 (1977).
- [6] D. Ruelle, *Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics*, 2nd ed. (Cambridge Univ. Press, Cambridge UK, 2004).
- [7] H. H. Rugh, “Un interview de David Ruelle”, *Gaz. Math.* **161**, 55–59 (2019).
- [8] H. H. Rugh, “An interview with David Ruelle”, *EMS Newsletter* **2020-3**, 21–24 (2020).
- [9] C. Shannon, “A mathematical theory of communication”, *Bell System Technical Journal* **27**, 379 (1948).
- [10] E. B. Vul, Y. G. Sinai, and K. M. Khanin, “Feigenbaum universality and the thermodynamic formalism”, *Russ. Math. Surv.* **39**, 1–40S (1984).

A32.3 Exercises

- A32.1. **Thermodynamics in higher dimensions.** Define Lyapunov exponents as the time averages of the eigen-exponents of the Jacobian matrix J

$$\mu^{(k)} = \lim_{t \rightarrow \infty} \frac{1}{t} \log |\Lambda_k^t(x_0)|, \quad (\text{A32.23})$$

as a generalization of (6.12).

Show that in d dimensions Pesin's formula is

$$K_1 = \sum_{k=1}^d \mu^{(k)} - \gamma, \quad (\text{A32.24})$$

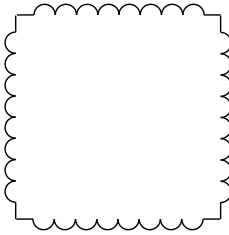
where the summation goes for the positive $\mu^{(k)}$ -s only. Hint: Use the d -dimensional generalization of (A32.9)

$$\mu_p = e^{n\gamma} / \left| \prod_k \Lambda_{p,k} \right|,$$

where the product goes for the expanding eigenvalues of the Jacobian matrix of p -cycle. (G. Vattay)

- A32.2. **Stadium billiard Kolmogorov entropy.** (continuation of exercise 9.6) Take $a = 1.6$ and $d = 1$ in figure 9.1, and estimate the Lyapunov exponent by averaging over a very long trajectory. Biham and Kvale [2] estimate the discrete time Lyapunov to $\lambda \approx 1.0 \pm .1$, the continuous time Lyapunov to $\lambda \approx 0.43 \pm .02$, the topological entropy (for their symbolic dynamics) $h \approx 1.15 \pm .03$.

- A32.3. **Entropy of rugged-edge billiards.** Take a semi-circle of diameter ε and replace the sides of a unit square by $\lfloor 1/\varepsilon \rfloor$ semi-circle arcs.



- (a) Is the billiard ergodic as $\varepsilon \rightarrow 0$?
 (b) (hard) Show that the entropy of the billiard map is

$$K_1 \rightarrow -\frac{2}{\pi} \ln \varepsilon + \text{const},$$

as $\varepsilon \rightarrow 0$. (Hint: do not write return maps.)

- (c) (harder) Show that when the semi-circles of the stadium billiard are far apart, say L , the entropy for the flow decays as

$$K_1 \rightarrow \frac{2 \ln L}{\pi L}.$$

- A32.4. **Two scale map.** Compute all those quantities - dimensions, escape rate, entropies, etc. - for the repeller of the one dimensional map

$$f(x) = \begin{cases} 1 + ax & \text{if } x < 0, \\ 1 - bx & \text{if } x > 0. \end{cases} \quad (\text{A32.25})$$

where a and b are larger than 2. Compute the fractal dimension, plot the pressure and compute the $f(\alpha)$ spectrum of singularities.

- A32.5. **Transfer matrix.** Take the unimodal map $f(x) = \sin(\pi x)$ of the interval $I = [0, 1]$. Calculate the four preimages of the intervals $I_0 = [0, 1/2]$ and $I_1 = [1/2, 1]$. Extrapolate $f(x)$ with piecewise linear functions on these intervals. Find a_1 , a_2 and b of the previous exercise. Calculate the pressure function of this linear extrapolation. Work out higher level approximations by linearly extrapolating the map on the 2^n -th preimages of I .

Appendix A33

Statistical mechanics recycled

(R. Mainieri)

A SPIN SYSTEM with long-range interactions can be converted into a chaotic dynamical system that is differentiable and low-dimensional. The thermodynamic limit quantities of the spin system are then equivalent to long time averages of the dynamical system. In this way the spin system averages can be recast as the cycle expansions. If the resulting dynamical system is analytic, the convergence to the thermodynamic limit is faster than with the standard transfer matrix techniques.

A33.1 The thermodynamic limit

There are two motivations to recycle statistical mechanics: one gets better control over the thermodynamic limit and one gets detailed information on how one is converging to it. From this information, most other quantities of physical interest can be computed.

In statistical mechanics one computes the averages of observables. These are functions that return a number for every state of the system; they are an abstraction of the process of measuring the pressure or temperature of a gas. The average of an observable is computed in the thermodynamic limit — the limit of system with an arbitrarily large number of particles. The thermodynamic limit is an essential step in the computation of averages, as it is only then that one observes the bulk properties of matter.

Without the thermodynamic limit many of the thermodynamic properties of matter could not be derived within the framework of statistical mechanics. There would be no extensive quantities, no equivalence of ensembles, and no phase transitions. From experiments it is known that certain quantities are extensive, that is,

they are proportional to the size of the system. This is not true for an interacting set of particles. If two systems interacting via pairwise potentials are brought close together, work will be required to join them, and the final total energy will not be the sum of the energies of each of the parts. To avoid the conflict between the experiments and the theory of Hamiltonian systems, one needs systems with an infinite number of particles. In the canonical ensemble the probability of a state is given by the Boltzmann factor which does not impose the conservation of energy; in the microcanonical ensemble energy is conserved but the Boltzmann factor is no longer exact. The equality between the ensembles only appears in the limit of the number of particles going to infinity at constant density. The phase transitions are interpreted as points of non-analyticity of the free energy in the thermodynamic limit. For a finite system the partition function cannot have a zero as a function of the inverse temperature β , as it is a finite sum of positive terms.

The thermodynamic limit is also of central importance in the study of field theories. A field theory can be first defined on a lattice and then the lattice spacing is taken to zero as the correlation length is kept fixed. This continuum limit corresponds to the thermodynamic limit. In lattice spacing units the correlation length is going to infinity, and the interacting field theory can be thought of as a statistical mechanics model at a phase transition.

For general systems the convergence towards the thermodynamic limit is slow. If the thermodynamic limit exists for an interaction, the convergence of the free energy per unit volume f is as an inverse power in the linear dimension of the system.

$$f(\beta) \rightarrow \frac{1}{n} \quad (\text{A33.1})$$

where n is proportional to $V^{1/d}$, with V the volume of the d -dimensional system. Much better results can be obtained if the system can be described by a transfer matrix. A transfer matrix is concocted so that the trace of its n th power is exactly the partition function of the system with one of the dimensions proportional to n . When the system is described by a transfer matrix then the convergence is exponential,

$$f(\beta) \rightarrow e^{-\alpha n} \quad (\text{A33.2})$$

and may only be faster than that if all long-range correlations of the system are zero — that is, when there are no interactions. The coefficient α depends only on the inverse correlation length of the system.

One of the difficulties in using the transfer matrix techniques is that they seem at first limited to systems with finite range interactions. Phase transitions can happen only when the interaction is long range. One can try to approximate the long range interaction with a series of finite range interactions that have an ever increasing range. The problem with this approach is that in a formally defined transfer matrix, not all the eigenvalues of the matrix correspond to eigenvalues of the system (in the sense that the rate of decay of correlations is not the ratio of eigenvalues).

Knowledge of the correlations used in conjunction with finite size scaling to obtain accurate estimates of the parameters of systems with phase transitions. (Accurate critical exponents are obtained by series expansions or transfer matrices, and infrequently by renormalization group arguments or Monte Carlo.) In a phase transition the coefficient α of the exponential convergence goes to zero and the convergence to the thermodynamic limit is power-law.

The computation of the partition function is an example of a functional integral. For most interactions these integrals are ill-defined and require some form of normalization. In the spin models case the functional integral is very simple, as “space” has only two points and only “time” being infinite has to be dealt with. The same problem occurs in the computation of the trace of transfer matrices of systems with infinite range interactions. If one tries to compute the partition function Z_n

$$Z_n = \text{tr } T^n$$

when T is an infinite matrix, the result may be infinite for any n . This is not to say that Z_n is infinite, but that the relation between the trace of an operator and the partition function breaks down. We could try regularizing the expression, but as we shall see below, that is not necessary, as there is a better physical solution to this problem.

What will be described here solves both of these problems in a limited context: it regularizes the transfer operator in a physically meaningful way, and as a consequence, it allows for the faster than exponential convergence to the thermodynamic limit and complete determination of the spectrum. The steps to achieve this are:

- Redefine the transfer operator so that there are no limits involved except for the thermodynamic limit.
- Note that the divergences of this operator come from the fact that it acts on a very large space. All that is needed is the smallest subspace containing the eigenvector corresponding to the largest eigenvalue (the Gibbs state).
- Rewrite all observables as depending on a local effective field. The eigenvector is like that, and the operator restricted to this space is trace-class.
- Compute the spectrum of the transfer operator and observe the magic.

A33.2 Ising models

The Ising model is a simple model to study the cooperative effects of many small interacting magnetic dipoles. The dipoles are placed on a lattice and their interaction is greatly simplified. There can also be a field that includes the effects of an external magnetic field and the average effect of the dipoles among themselves. We will define a general class of Ising models (also called spin systems) where the

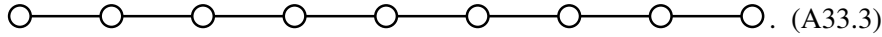
dipoles can be in one of many possible states and the interactions extend beyond the nearest neighboring sites of the lattice. But before we extend the Ising model, we will examine the simplest model in that class.

A33.2.1 Ising model

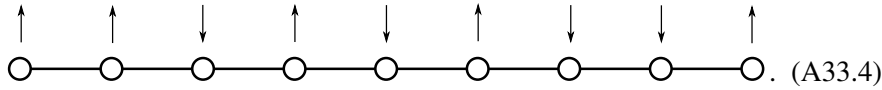
One of the simplest models in statistical mechanics is the Ising model. One imagines that one has a 1-dimensional lattice with small magnets at each site that can point either up or down.



Each little magnet interacts only with its neighbors. If they both point in the same direction, then they contribute an energy $-J$ to the total energy of the system; and if they point in opposite directions, then they contribute $+J$. The signs are chosen so that they prefer to be aligned. Let us suppose that we have n small magnets arranged in a line: A line is drawn between two sites to indicate that there is an interaction between the small magnets that are located on that site



(This figure can be thought of as a graph, with sites being vertices and interacting magnets indicated by edges.) To each of the sites we associate a variable, that we call a spin, that can be in either of two states: up ($\uparrow$) or down ($\downarrow$). This represents the two states of the small magnet on that site, and in general we will use the notation Σ_0 to represent the set of possible values of a spin at any site; all sites assume the same set of values. A configuration consists of assigning a value to the spin at each site; a typical configuration is



The set of all configurations for a lattice with n sites is called Ω_0^n and is formed by the Cartesian product $\Omega_0 \times \Omega_0 \cdots \times \Omega_0$, the product repeated n times. Each configuration $\sigma \in \Omega^n$ is a string of n spins

$$\sigma = \{\sigma_0, \sigma_1, \dots, \sigma_n\}, \quad (\text{A33.5})$$

In the example configuration (A33.4) there are two pairs of spins that have the same orientation and six that have the opposite orientation. Therefore the total energy H of the configuration is $J \times 6 - J \times 2 = 4J$. In general we can associate an energy H to every configuration

$$H(\sigma) = \sum_i J\delta(\sigma_i, \sigma_{i+1}), \quad (\text{A33.6})$$

where

$$\delta(\sigma_1, \sigma_2) = \begin{cases} +1 & \text{if } \sigma_1 = \sigma_2 \\ -1 & \text{if } \sigma_1 \neq \sigma_2 \end{cases}. \quad (\text{A33.7})$$

One of the problems that was avoided when computing the energy was what to do at the boundaries of the 1-dimensional chain. Note that as written, (A33.6) requires the interaction of spin n with spin $n + 1$. In the absence of phase transitions the boundaries do not matter much to the thermodynamic limit and we will connect the first site to the last, implementing periodic boundary conditions.

Thermodynamic quantities are computed from the partition function $Z^{(n)}$ as the size n of the system becomes very large. For example, the free energy per site f at inverse temperature β is given by

$$-\beta f(\beta) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln Z^{(n)}. \quad (\text{A33.8})$$

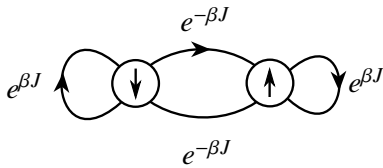
The partition function $Z^{(n)}$ is computed by a sum that runs over all the possible configurations on the 1-dimensional chain. Each configuration contributes with its Gibbs factor $\exp(-\beta H(\sigma))$ and the partition function $Z^{(n)}$ is

$$Z^{(n)}(\beta) = \sum_{\sigma \in \Omega_0^n} e^{-\beta H(\sigma)}. \quad (\text{A33.9})$$

The partition function can be computed using transfer matrices. This is a method that generalizes to other models. At first, it is a little mysterious that matrices show up in the study of a sum. To see where they come from, we can try and build a configuration on the lattice site by site. The first thing to do is to expand out the sum for the energy of the configuration

$$Z^{(n)}(\beta) = \sum_{\sigma \in \Omega^n} e^{\beta J \delta(\sigma_1, \sigma_2)} e^{\beta J \delta(\sigma_2, \sigma_3)} \dots e^{\beta J \delta(\sigma_n, \sigma_1)}. \quad (\text{A33.10})$$

Let us use the configuration in (A33.4). The first site is $\sigma_1 = \uparrow$. As the second site is $\uparrow$, we know that the first term in (A33.10) is a term $e^{\beta J}$. The third spin is $\downarrow$, so the second term in (A33.10) is $e^{-\beta J}$. If the third spin had been $\uparrow$, then the term would have been $e^{\beta J}$ but it would not depend on the value of the first spin σ_1 . This means that the configuration can be built site by site and that to compute the Gibbs factor for the configuration just requires knowing the last spin added. We can then think of the configuration as being a weighted random walk where each step of the walk contributes according to the last spin added. The random walk take place on the transition graph



Choose one of the two sites as a starting point. Walk along any allowed edge making your choices randomly and keep track of the accumulated weight as you perform the n steps. To implement the periodic boundary conditions make sure that you return to the starting node of the transition graph. If the walk is carried out in all possible 2^n ways then the sum of all the weights is the partition function. To perform the sum we consider the matrix

$$T(\beta) = \begin{bmatrix} e^{\beta J} & e^{-\beta J} \\ e^{-\beta J} & e^{\beta J} \end{bmatrix}. \quad (\text{A33.11})$$

As in chapter 14 the sum of all closed walks is given by the trace of powers of the matrix. These powers can easily be re-expressed in terms of the two eigenvalues λ_1 and λ_2 of the transfer matrix:

$$Z^{(n)}(\beta) = \text{tr } T^n(\beta) = \lambda_1(\beta)^n + \lambda_2(\beta)^n. \quad (\text{A33.12})$$

A33.2.2 Averages of observables

Averages of observables can be re-expressed in terms of the eigenvectors of the transfer matrix. Alternatively, one can introduce a modified transfer matrix and compute the averages through derivatives. Sounds familiar?

A33.2.3 General spin models

The more general version of the Ising model — the spin models — will be defined on a regular lattice, $\mathbb{Z}^D$. At each lattice site there will be a spin variable that can assume a finite number of states identified by the set Ω_0 .

The transfer operator $\mathcal{T}$ was introduced by Kramers and Wannier [11] to study the Ising model on a strip and concocted so that the trace of its n th power is the partition function Z_n of system when one of its dimensions is n . The method can be generalized to deal with any finite-range interaction. If the range of the interaction is L , then $\mathcal{T}$ is a matrix of size $2^L \times 2^L$. The longer the range, the larger the matrix.

A33.3 Fisher droplet model

In a series of articles [7], Fisher introduced the droplet model. It is a model for a system containing two phases: gas and liquid. At high temperatures, the typical state of the system consists of droplets of all sizes floating in the gas phase. As the temperature is lowered, the droplets coalesce, forming larger droplets, until at the transition temperature, all droplets form one large one. This is a first order phase transition.

Although Fisher formulated the model in 3-dimensions, the analytic solution of the model shows that it is equivalent to a 1-dimensional lattice gas model with long range interactions. Here we will show how the model can be solved for an arbitrary interaction, as the solution only depends on the asymptotic behavior of the interaction.

The interest of the model for the study of cycle expansions is its relation to intermittency. By having an interaction that behaves asymptotically as the scaling function for intermittency, one expects that the analytic structure (poles and cuts) will be same.

Fisher used the droplet model to study a first order phase transition [7]. Gallavotti [8] used it to show that the zeta functions cannot in general be extended to a meromorphic functions of the entire complex plane. The droplet model has also been used in dynamical systems to explain features of mode locking, see Artuso [1]. In computing the zeta function for the droplet model we will discover that at low temperatures the cycle expansion has a limited radius of convergence, but it is possible to factorize the expansion into the product of two functions, each of them with a better understood radius of convergence.

A33.3.1 Solution

The droplet model is a 1-dimensional lattice gas where each site can have two states: empty or occupied. We will represent the empty state by 0 and the occupied state by 1. The configurations of the model in this notation are then strings of zeros and ones. Each configuration can be viewed as groups of contiguous ones separated by one or more zeros. The contiguous ones represent the droplets in the model. The droplets do not interact with each other, but the individual particles within each droplet do.

To determine the thermodynamics of the system we must assign an energy to every configuration. At very high temperatures we would expect a gaseous phase where there are many small droplets, and as we decrease the temperature the droplets would be expected to coalesce into larger ones until at some point there is a phase transition and the configuration is dominated by one large drop. To construct a solvable model and yet one with a phase transition we need long range interaction among all the particles of a droplet. One choice is to assign a fixed energy θ_n for the interactions of the particles of a cluster of size n . In a given droplet one has to consider all the possible clusters formed by contiguous particles. Consider for example the configuration 0111010. It has two droplets, one of size three and another of size one. The droplet of size one has only one cluster of size one and therefore contributes to the energy of the configuration with θ_1 . The cluster of size three has one cluster of size three, two clusters of size two, and three clusters of size one; each cluster contributing a θ_n term to the energy. The total energy of the configuration is then

$$H(0111010) = 4\theta_1 + 2\theta_2 + 1\theta_3. \quad (\text{A33.13})$$

If there were more zeros around the droplets in the above configuration the energy would still be the same. The interaction of one site with the others is assumed to be finite, even in the ground state consisting of a single droplet, so there is a restriction on the sum of the cluster energies given by

$$a = \sum_{n>0} \theta_n < \infty. \quad (\text{A33.14})$$

The configuration with all zeros does not contribute to the energy.

Once we specify the function θ_n we can compute the energy of any configuration, and from that determine the thermodynamics. Here we will evaluate the

cycle expansion for the model by first computing the generating function

$$G(z, \beta) = \sum_{n>0} z^n \frac{Z_n(\beta)}{n} \quad (\text{A33.15})$$

and then considering its exponential, the cycle expansion. Each partition function Z_n must be evaluated with periodic boundary conditions. So if we were computing Z_3 we must consider all eight binary sequences of three bits, and when computing the energy of a configuration, say 011, we should determine the energy per three sites of the long chain

$$\dots 011011011011 \dots$$

In this case the energy would be $\theta_2 + 2\theta_1$. If instead of 011 we had considered one of its rotated shifts, 110 or 101, the energy of the configuration would have been the same. To compute the partition function we only need to consider one of the configurations and multiply by the length of the configuration to obtain the contribution of all its rotated shifts. The factor $1/n$ in the generating function cancels this multiplicative factor. This reduction will not hold if the configuration has a symmetry, as for example 0101 which has only two rotated shift configurations. To compensate this we replace the $1/n$ factor by a symmetry factor $1/s(b)$ for each configuration b . The evaluation of G is now reduced to summing over all configurations that are not rotated shift equivalent, and we call these the basic configurations and the set of all of them B . We now need to evaluate

$$G(z, \beta) = \sum_{b \in B} \frac{z^{|b|}}{s(b)} e^{-\beta H(b)}. \quad (\text{A33.16})$$

The notation $|\cdot|$ represents the cardinality of the set.

Any basic configuration can be built by considering the set of droplets that form it. The smallest building block has size two, as we must also put a zero next to the one so that when two different blocks get put next to each other they do not coalesce. The first few building blocks are

size	droplets	
2	01	(A33.17)
3	001 011	
4	0001 0011 0111	

Each droplet of size n contributes with energy

$$W_n = \sum_{1 \leq k \leq n} (n - k + 1) \theta_k. \quad (\text{A33.18})$$

So if we consider the sum

$$\sum_{n \geq 1} \frac{1}{n} \left(z^2 e^{-\beta H(01)} + z^3 (e^{-\beta H(001)} + e^{-\beta H(011)}) + \right. \\ \left. + z^4 (e^{-\beta H(0001)} + e^{-\beta H(0011)} + e^{-\beta H(0111)}) + \dots \right)^n \quad (\text{A33.19})$$

then the power in n will generate all the configurations that are made from many droplets, while the z will keep track of the size of the configuration. The factor $1/n$ is there to avoid the over-counting, as we only want the basic configurations and not its rotated shifts. The $1/n$ factor also gives the correct symmetry factor in the case the configuration has a symmetry. The sum can be simplified by noticing that it is a logarithmic series

$$-\ln\left(1 - (z^2 e^{-\beta W_1} + z^3(e^{-\beta W_1} + e^{-\beta W_2}) + \dots)\right), \quad (\text{A33.20})$$

where the $H(b)$ factors have been evaluated in terms of the droplet energies W_n . A proof of the equality of (A33.19) and (A33.20) can be given, but we there was not enough space on the margin to write it down. The series that is subtracted from one can be written as a product of two series and the logarithm written as

$$-\ln\left(1 - (z^1 + z^2 + z^3 + \dots)(ze^{-\beta W_1} + z^2 e^{-\beta W_2} + \dots)\right) \quad (\text{A33.21})$$

The product of the two series can be directly interpreted as the generating function for sequences of droplets. The first series adds one or more zeros to a configuration and the second series add a droplet.

There is a whole class of configurations that is not included in the above sum: the configurations formed from a single droplet and the vacuum configuration. The vacuum is the easiest, as it has zero energy it only contributes a z . The sum of all the null configurations of all sizes is

$$\sum_{n>0} \frac{z^n}{n}. \quad (\text{A33.22})$$

The factor $1/n$ is here because the original G had them and the null configurations have no rotated shifts. The single droplet configurations also do not have rotated shifts so their sum is

$$\sum_{n>0} \frac{z^n e^{-\beta H(\overbrace{11 \dots 11}^n)}}{n}. \quad (\text{A33.23})$$

Because there are no zeros in the above configuration clusters of all size exist and the energy of the configuration is $n \sum \theta_k$ which we denote by na .

From the three sums (A33.21), (A33.22), and (A33.23) we can evaluate the generating function G to be

$$G(z, \beta) = -\ln(1 - z) - \ln(1 - ze^{-\beta a}) - \ln\left(1 - \frac{z}{1 - z} \sum_{n \geq 1} z^n e^{-\beta W_n}\right). \quad (\text{A33.24})$$

The cycle expansion $\zeta^{-1}(z, \beta)$ is given by the exponential of the generating function e^{-G} and we obtain

$$\zeta^{-1}(z, \beta) = (1 - ze^{-\beta a})(1 - z(1 + \sum_{n \geq 1} z^n e^{-\beta W_n})) \quad (\text{A33.25})$$

To pursue this model further we need to have some assumptions about the interaction strengths θ_n . We will assume that the interaction strength decreases with the inverse square of the size of the cluster, that is, $\theta_n = -1/n^2$. With this we can estimate that the energy of a droplet of size n is asymptotically

$$W_n \sim -n + \ln n + O\left(\frac{1}{n}\right). \quad (\text{A33.26})$$

If the power chosen for the polynomially decaying interaction had been other than inverse square we would still have the droplet term proportional to n , but there would be no logarithmic term, and the O term would be of a different power. The term proportional to n survives even if the interactions falls off exponentially, and in this case the correction is exponentially small in the asymptotic formula. To simplify the calculations we are going to assume that the droplet energies are exactly

$$W_n = -n + \ln n \quad (\text{A33.27})$$

in a system of units where the dimensional constants are one. To evaluate the cycle expansion (A33.25) we need to evaluate the constant a , the sum of all the θ_n . One can write a recursion for the θ_n

$$\theta_n = W_n - \sum_{1 \leq k < n} (n - k + 1) \theta_k \quad (\text{A33.28})$$

and with an initial choice for θ_1 evaluate all the others. It can be verified that independent of the choice of θ_1 the constant a is equal to the number that multiplies the n term in (A33.27). In the units used

$$a = -1. \quad (\text{A33.29})$$

For the choice of droplet energy (A33.27) the sum in the cycle expansion can be expressed in terms of a special function: the Lerch transcendental ϕ_L . It is defined by

$$\phi_L(z, s, c) = \sum_{n \geq 0} \frac{z^n}{(n + c)^s}, \quad (\text{A33.30})$$

excluding from the sum any term that has a zero denominator. The Lerch function converges for $|z| < 1$. The series can be analytically continued to the complex plane and it will have a branch point at $z = 1$ with a cut chosen along the positive real axis. In terms of Lerch transcendental function we can write the cycle expansion (A33.25) using (A33.27) as

$$\zeta^{-1}(z, \beta) = (1 - ze^\beta) \left(1 - z(1 + \phi_L(ze^\beta, \beta, 1)) \right) \quad (\text{A33.31})$$

This serves as an example of a zeta function that cannot be extended to a meromorphic function of the complex plane as one could conjecture.

The thermodynamics for the droplet model comes from the smallest root of (A33.31). The root can come from any of the two factors. For large value of β (low

temperatures) the smallest root is determined from the $(1 - ze^\beta)$ factor, which gave the contribution of a single large drop. For small β (large temperatures) the root is determined by the zero of the other factor, and it corresponds to the contribution from the gas phase of the droplet model. The transition occurs when the smallest root of each of the factors become numerically equal. This determines the critical temperature β_c through the equation

$$1 - e^{-\beta_c}(1 + \zeta_R(\beta_c)) = 0 \quad (\text{A33.32})$$

which can be solved numerically. One finds that $\beta_c = 1.40495$. The phase transition occurs because the roots from two different factors get swapped in their roles as the smallest root. This in general leads to a first order phase transition. For large β the Lerch transcendental is being evaluated at the branch point, and therefore the cycle expansion cannot be an analytic function at low temperatures. For large temperatures the smallest root is within the radius of convergence of the series for the Lerch transcendental, and the cycle expansion has a domain of analyticity containing the smallest root.

As we approach the phase transition point as a function of β the smallest root and the branch point get closer together until at exactly the phase transition they collide. This is a sufficient condition for the existence of a first order phase transitions. In the literature of zeta functions [12] there have been speculations on how to characterize a phase transition within the formalism. The solution of the Fisher droplet model suggests that for first order phase transitions the factorized cycle expansion will have its smallest root within the radius of convergence of one of the series except at the phase transition when the root collides with a singularity. This does not seem to be the case for second order phase transitions.

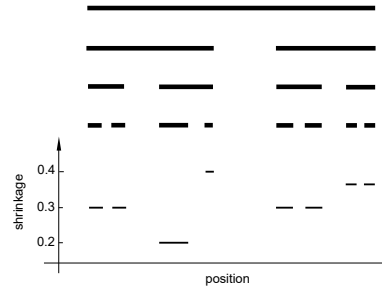
The analyticity of the cycle expansion can be restored if we consider separate cycle expansions for each of the phases of the system. If we separate the two terms of ζ^{-1} in (A33.31), each of them is an analytic function and contains the smallest root within the radius of convergence of the series for the relevant β values.

A33.4 Scaling functions

There is a relation between general spin models and dynamical system. If one thinks of the boxes of the Markov partition of a hyperbolic system as the states of a spin system, then computing averages in the dynamical system is carrying out a sum over all possible states. One can even construct the natural measure of the dynamical system from a translational invariant “interaction function” call the scaling function.

There are many routes that lead to an explanation of what a scaling function is and how to compute it. The shortest is by breaking away from the historical development and considering first the presentation function of a fractal. The presentation function is a simple chaotic dynamical system (hyperbolic, unlike the circle map) that generates the fractal and is closely related to the definition

Figure A33.1: Construction of the steps of the scaling function from a Cantor set. From one level to the next in the construction of the Cantor set, each parent segment changes into two children segments. The shrinkage of the last level of the construction is plotted and by removing the gaps one has an approximation to the scaling function of the Cantor set.



of fractals of Hutchinson [10] and the iterated dynamical systems introduced by Barnsley and collaborators [3]. From the presentation function one can derive the scaling function, but we will not do it in the most elegant fashion, rather we will develop the formalism in a form that is directly applicable to the experimental data.

In the upper part of figure A33.1 we have the successive steps of the construction similar to the middle third Cantor set. The construction is done in levels, each level being formed by a collection of segments. From one level to the next, each “parent” segment produces smaller “children” segments by removing the middle section. As the construction proceeds, the segments better approximate the Cantor set. In the figure not all the segments are the same size, some are larger and some are smaller, as is the generic case for a fractal (sometimes referred to as a ‘multifractal’). In the middle third Cantor set, the ratio between a segment and the one it was generated from is exactly $1/3$, but in the case shown in the figure the ratios differ from $1/3$. If we went through the last level of the construction and made a plot of the segment number and its ratio to its parent segment we would have a scaling function, as indicated in the figure. A function giving the ratios in the construction of a fractal is the basic idea for a scaling function. Much of the formalism that we will introduce is to be able to give precise names to every segments and to arrange the “lineage” of segments so that the children segments have the correct parent. If we do not take these precautions, the scaling function would be a “wild function,” varying rapidly and not approximated easily by simple functions.

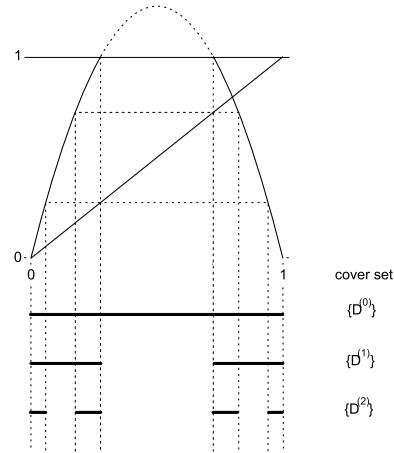
To describe the formalism we will use a variation on the quadratic map that appears in the theory of period doubling. This is because the combinatorial manipulations are much simpler for this map than they are for the circle map. The scaling function will be described for a one dimensional map F as shown in figure A33.2. Drawn is the map

$$F(x) = 5x(1 - x) \quad (\text{A33.33})$$

restricted to the unit interval. We will see that this map is also a presentation function.

It has two branches separated by a gap: one over the left portion of the unit interval and one over the right. If we choose a point x at random in the unit interval and iterate it under the action of the map F , (A33.33), it will hop between the branches and eventually get mapped to minus infinity. An orbit point is guaranteed

Figure A33.2: A Cantor set presentation function. The Cantor set is the set of all points that under iteration do not leave the interval $[0, 1]$. This set can be found by backwards iterating the gap between the two branches of the map. The dotted lines can be used to find these backward images. At each step of construction, a set of segments exists that forms a cover for the Cantor set.



to go to minus infinity if it lands in the gap. The hopping of the point defines the orbit of the initial point x : $x \mapsto x_1 \mapsto x_2 \mapsto \dots$. For each orbit of the map F we can associate a symbolic code. The code for this map is formed from 0s and 1s and is found from the orbit by associating a 0 if $x_t < 1/2$ and a 1 if $x_t > 1/2$, with $t = 0, 1, 2, \dots$

Most initial points will end up in the gap region between the two branches. We then say that the orbit point has escaped the unit interval. The points that do not escape form a Cantor set C (or Cantor dust) and remain trapped in the unit interval for all iterations. In the process of describing all the points that do not escape, the map F can be used as a presentation of the Cantor set C , and has been called a presentation function by Feigenbaum [6].

How does the map F “present” the Cantor set? The presentation is done in steps. First, we determine the points that do not escape the unit interval in one iteration of the map. These are the points that are not part of the gap. These points determine two segments, which are an approximation to the Cantor set. In the next step we determine the points that do not escape in two iterations. These are the points that get mapped into the gap in one iteration, as in the next iteration they will escape; these points form the two segments $\Delta_0^{(1)}$ and $\Delta_1^{(1)}$ at level 1 in figure A33.2. The processes can be continued for any number of iterations. If we observe carefully what is being done, we discover that at each step the pre-images of the gap (backward iterates) are being removed from the unit interval. As the map has two branches, every point in the gap has two pre-images, and therefore the whole gap has two pre-images in the form of two smaller gaps. To generate all the gaps in the Cantor set one just has to iterate the gap backwards. Each iteration of the gap defines a set of segments, with the n th iterate defining the segments $\Delta_k^{(n)}$ at level n . For this map there will be 2^n segments at level n , with the first few drawn in figure A33.2. As $n \rightarrow \infty$ the segments that remain for at least n iterates converge to the Cantor set C .

The segments at one level form a cover for the Cantor set and it is from a cover that all the invariant information about the set is extracted (the cover generated from the backward iterates of the gap form a Markov partition for the map as a dynamical system). The segments $\{\Delta_k^{(n)}\}$ at level n are a refinement of the cover

formed by segments at level $n - 1$. From successive covers we can compute the trajectory scaling function, the spectrum of scalings $f(\alpha)$, and the generalized dimensions.

To define the scaling function we must give labels (names) to the segments. The labels are chosen so that the definition of the scaling function allows for simple approximations. As each segment is generated from an inverse image of the unit interval, we will consider the inverse of the presentation function F . Because F does not have a unique inverse, we have to consider restrictions of F . Its restriction to the first half of the segment, from 0 to $1/2$, has a unique inverse, which we will call F_0^{-1} , and its restriction to the second half, from $1/2$ to 1, also has a unique inverse, which we will call F_1^{-1} . For example, the segment labeled $\Delta^{(2)}(0, 1)$ in figure A33.2 is formed from the inverse image of the unit interval by mapping $\Delta^{(0)}$, the unit interval, with F_1^{-1} and then F_0^{-1} , so that the segment

$$\Delta^{(2)}(0, 1) = F_0^{-1} \left(F_1^{-1} \left(\Delta^{(0)} \right) \right). \quad (\text{A33.34})$$

The mapping of the unit interval into a smaller interval is what determines its label. The sequence of the labels of the inverse maps is the label of the segment:

$$\Delta^{(n)}(\epsilon_1, \epsilon_2, \dots, \epsilon_n) = F_{\epsilon_1}^{-1} \circ F_{\epsilon_2}^{-1} \circ \dots \circ F_{\epsilon_n}^{-1} \left(\Delta^{(0)} \right).$$

The scaling function is formed from a set of ratios of segments length. We use $|\cdot|$ around a segment $\Delta^{(n)}(\epsilon)$ to denote its size (length), and define

$$\sigma^{(n)}(\epsilon_1, \epsilon_2, \dots, \epsilon_n) = \frac{|\Delta^{(n)}(\epsilon_1, \epsilon_2, \dots, \epsilon_n)|}{|\Delta^{(n-1)}(\epsilon_2, \dots, \epsilon_n)|}.$$

We can then arrange the ratios $\sigma^{(n)}(\epsilon_1, \epsilon_2, \dots, \epsilon_n)$ next to each other as piecewise constant segments in increasing order of their binary label $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ so that the collection of steps scan the unit interval. As $n \rightarrow \infty$ this collection of steps will converge to the scaling function.

A33.5 Geometrization

The $\mathcal{L}$ operator is a generalization of the transfer matrix. It gets more by considering less of the matrix: instead of considering the whole matrix it is possible to consider just one of the rows of the matrix. The $\mathcal{L}$ operator also makes explicit the vector space in which it acts: that of the observable functions. Observables are functions that to each configuration of the system associate a number: the energy, the average magnetization, the correlation between two sites. It is in the average of observables that one is interested in. Like the transfer matrix, the $\mathcal{L}$ operator considers only semi-infinite systems, that is, only the part of the interaction between spins to the right is taken into account. This may sound un-symmetric, but it is a simple way to count each interaction only once, even in cases where the interaction includes three or more spin couplings. To define the $\mathcal{L}$ operator one

needs the interaction energy between one spin and all the rest to its right, which is given by the function ϕ . The $\mathcal{L}$ operators defined as

$$\mathcal{L}g(\sigma) = \sum_{\sigma_0 \in \Omega_0} g(\sigma_0 \sigma) e^{-\beta \phi(\sigma_0 \sigma)}.$$

To each possible value in Ω_0 that the spin σ_0 can assume, an average of the observable g is computed weighed by the Boltzmann factor $e^{-\beta \phi}$. The formal relations that stem from this definition are its relation to the free energy when applied to the observable ι that returns one for any configuration:

$$-\beta f(\beta) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \|\mathcal{L}^n \iota\|$$

and the thermodynamic average of an observable

$$\langle g \rangle = \lim_{n \rightarrow \infty} \frac{\|\mathcal{L}^n g\|}{\|\mathcal{L}^n \iota\|}.$$

Both relations hold for almost all configurations. These relations are part of the theorem of Ruelle that enlarges the domain of the Perron-Frobenius theorem and sharpens its results. The theorem shows that just as the transfer matrix, the largest eigenvalue of the $\mathcal{L}$ operator is related to the free-energy of the spin system. It also shows that there is a formula for the eigenvector related to the largest eigenvalue. This eigenvector $|\rho\rangle$ (or the corresponding one for the adjoint $\mathcal{L}^*$ of $\mathcal{L}$) is the Gibbs state of the system. From it all averages of interest in statistical mechanics can be computed from the formula

$$\langle g \rangle = \langle \rho | g | \rho \rangle.$$

The Gibbs state can be expressed in an explicit form in terms of the interactions, but it is of little computational value as it involves the Gibbs state for a related spin system. Even then it does have an enormous theoretical value. Later we will see how the formula can be used to manipulate the space of observables into a more convenient space.

The geometrization of a spin system converts the shift dynamics (necessary to define the Ruelle operator) into a smooth dynamics. This is equivalent to the mathematical problem in ergodic theory of finding a smooth embedding for a given Bernoulli map.

The basic idea for the dynamics is to establish the a set of maps F_{σ_k} such that

$$F_{\sigma_k}(0) = 0$$

and

$$F_{\sigma_1} \circ F_{\sigma_2} \circ \cdots \circ F_{\sigma_n}(0) = \phi(+, \sigma_1, \sigma_2, \dots, \sigma_n, -, -, \dots).$$

This is a formal relation that expresses how the interaction is to be converted into a dynamical systems. In most examples F_{σ_k} is a collection of maps from a subset of R^D to itself.

If the interaction is complicated, then the dimension of the set of maps may be infinite. If the resulting dynamical system is infinite have we gained anything from the transformation? The gain in this case is not in terms of added speed of convergence to the thermodynamic limit, but in the fact that the Ruelle operator is of trace-class and all eigenvalues are related to the spin system and not artifacts of the computation.

The construction of the higher dimensional system is done by borrowing the state space reconstruction technique from dynamical systems. State space reconstruction can be done in several ways: by using delay coordinates, by using derivatives of the position, or by considering the value of several independent observables of the system. All these may be used in the construction of the equivalent dynamics. Just as in the study of dynamical systems, the exact method does not matter for the determination of the thermodynamics ($f(\alpha)$ spectra, generalized dimension), also in the construction of the equivalent dynamics the exact choice of observable does not matter.

We will only consider configurations for the half line. This is because for translational invariant interactions the thermodynamic limit on half line is the same as in the whole line. One can prove this by considering the difference in a thermodynamic average in the line and in the semiline and compare the two as the size of the system goes to infinity.

When the interactions are long range in principle one has to specify the boundary conditions to be able to compute the interaction energy of a configuration in a finite box. If there are no phase transitions for the interaction, then which boundary conditions are chosen is irrelevant in the thermodynamic limit. When computing quantities with the transfer matrix, the long range interaction is truncated at some finite range and the truncated interaction is then used to evaluate the transfer matrix. With the Ruelle operator the interaction is never truncated, and the boundary must be specified.

The interaction $\phi(\sigma)$ is any function that returns a number on a configuration. In general it is formed from pairwise spin interactions

$$\phi(\sigma) = \sum_{n>0} \delta_{\sigma_0, \sigma_n} J(n)$$

with different choices of $J(n)$ leading to different models. If $J(n) = 1$ only if $n = 1$ and 0 otherwise, then one has the nearest neighbor Ising model. If $J(n) = n^{-2}$, then one has the inverse square model relevant in the study of the Kondo problem.

Let us say that each site of the lattice can assume two values $+$, $-$ and the set of all possible configurations of the semiline is the set Ω . Then an observable g is a function from the set of configurations Ω to the reals. Each configuration is indexed by the integers from 0 up, and it is useful to think of the configuration as a string of spins. One can append a spin η_0 to its beginning, $\eta \vee \sigma$, in which case η is at site 0, ω_0 at site 1, and so on.

The Ruelle operator $\mathcal{L}$ is defined as

$$\mathcal{L}g(\eta) = \sum_{\omega_0 \in \Omega_0} g(\omega_0 \vee \eta) e^{-\beta\phi(\omega_0 \vee \eta)}.$$

This is a positive and bounded operator over the space of bounded observables. There is a generalization of the Perron-Frobenius theorem by Ruelle that establishes that the largest eigenvalue of $\mathcal{L}$ is isolated from the rest of the spectrum and gives the thermodynamics of the spin system just as the largest eigenvalue of the transfer matrix does. Ruelle also gave a formula for the eigenvector related to the largest eigenvalue.

The difficulty with it is that the relation between the partition function and the trace of its n th power, $\text{tr } \mathcal{L}^n = Z_n$ no longer holds. The reason is that the trace of the Ruelle operator is ill-defined, it is infinite.

We now introduce a special set of observables $\{x_1(\sigma), \dots, x_1(\sigma)\}$. The idea is to choose the observables in such a way that from their values on a particular configuration σ the configuration can be reconstructed. We also introduce the interaction observables h_{σ_0} .

To geometrize spin systems, the interactions are assumed to be translationally invariant. The spins σ_k will only assume a finite number of values. For simplicity, we will take the interaction ϕ among the spins to depend only on pairwise interactions,

$$\phi(\sigma) = \phi(\sigma_0, \sigma_1, \sigma_2, \dots) = J_0 \sigma_0 + \sum_{n>0} \delta_{\sigma_0, \sigma_n} J_1(n), \quad (\text{A33.35})$$

and limit σ_k to be in $\{+, -\}$. For the 1-dimensional Ising model, J_0 is the external magnetic field and $J_1(n) = 1$ if $n = 1$ and 0 otherwise. For an exponentially decaying interaction $J_1(n) = e^{-an}$. Two- and 3-dimensional models can be considered in this framework. For example, a strip of spins of $L \times \infty$ with helical boundary conditions is modeled by the potential $J_1(n) = \delta_{n,1} + \delta_{n,L}$.

The transfer operator $\mathcal{T}$ was introduced by Kramers and Wannier [11] to study the Ising model on a strip and concocted so that the trace of its n th power is the partition function Z_n of system when one of its dimensions is n . The method can be generalized to deal with any finite-range interaction. If the range of the interaction is L , then $\mathcal{T}$ is a matrix of size $2^L \times 2^L$. The longer the range, the larger the matrix. When the range of the interaction is infinite one has to define the $\mathcal{T}$ operator by its action on an observable g . Just as the observables in quantum mechanics, g is a function that associates a number to every state (configuration of spins). The energy density and the average magnetization are examples of observables. From this equivalent definition one can recover the usual transfer matrix by making all quantities finite range. For a semi-infinite configuration $\sigma = \{\sigma_0, \sigma_1, \dots\}$:

$$\mathcal{T}g(\sigma) = g(+ \vee \sigma) e^{-\beta\phi(+ \vee \sigma)} + g(- \vee \sigma) e^{-\beta\phi(- \vee \sigma)}. \quad (\text{A33.36})$$

By $+ \vee \sigma$ we mean the configuration obtained by prepending $+$ to the beginning of σ resulting in the configuration $\{+, \sigma_0, \sigma_1, \dots\}$. When the range becomes infinite, $\text{tr } \mathcal{T}^n$ is infinite and there is no longer a connection between the trace and

the partition function for a system of size n (this is a case where matrices give the wrong intuition). Ruelle [14] generalized the Perron-Frobenius theorem and showed that even in the case of infinite range interactions the largest eigenvalue of the $\mathcal{T}$ operator is related to the free-energy of the spin system and the corresponding eigenvector is related to the Gibbs state. By applying $\mathcal{T}$ to the constant observable u , which returns 1 for any configuration, the free energy per site f is computed as

$$-\beta f(\beta) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \|\mathcal{T}^n u\|. \quad (\text{A33.37})$$

To construct a smooth dynamical system that reproduces the properties of $\mathcal{T}$, one uses the phase space reconstruction technique of Packard *et al.* [13] and Takens [17], and introduces a vector of state observables $x(\sigma) = \{x_1(\sigma), \dots, x_D(\sigma)\}$. To avoid complicated notation we will limit the discussion to the example $x(\sigma) = \{x_+(\sigma), x_-(\sigma)\}$, with $x_+(\sigma) = \phi(+ \vee \sigma)$ and $x_-(\sigma) = \phi(- \vee \sigma)$; the more general case is similar and used in a later example. The observables are restricted to those g for which, for all configurations σ , there exist an analytic function G such that $G(x_1(\sigma), \dots, x_D(\sigma)) = g(\sigma)$. This at first seems a severe restriction as it may exclude the eigenvector corresponding to the Gibbs state. It can be checked that this is not the case by using the formula given by Ruelle [15] for this eigenvector. A simple example where this formalism can be carried out is for the interaction $\phi(\sigma)$ with pairwise exponentially decaying potential $J_1(n) = a^n$ (with $|a| < 1$). In this case $\phi(\sigma) = \sum_{n>0} \delta_{\sigma_0, \sigma_n} a^n$ and the state observables are $x_+(\sigma) = \sum_{n>0} \delta_{+, \sigma_n} a^n$ and $x_-(\sigma) = \sum_{n>0} \delta_{-, \sigma_n} a^n$. In this case the observable x_+ gives the energy of $+$ spin at the origin, and x_- the energy of a $-$ spin.

Using the observables x_+ and x_- , the transfer operator can be re-expressed as

$$\mathcal{T}G(x(\sigma)) = \sum_{\eta \in \{+, -\}} G(x_+(\eta \vee \sigma), x_-(\eta \vee \sigma)) e^{-\beta x_\eta(\sigma)}. \quad (\text{A33.38})$$

In this equation the only reference to the configuration σ is when computing the new observable values $x_+(\eta \vee \sigma)$ and $x_-(\eta \vee \sigma)$. The iteration of the function that gives these values in terms of $x_+(\sigma)$ and $x_-(\sigma)$ is the dynamical system that will reproduce the properties of the spin system. For the simple exponentially decaying potential this is given by two maps, F_+ and F_- . The map F_+ takes $\{x_+(\sigma), x_-(\sigma)\}$ into $\{x_+(+ \vee \sigma), x_-(+ \vee \sigma)\}$ which is $\{a(1 + x_+), ax_-\}$ and the map F_- takes $\{x_+, x_-\}$ into $\{ax_+, a(1 + x_-)\}$. In a more general case we have maps F_η that take $x(\sigma)$ to $x(\eta \vee \sigma)$.

We can now define a new operator $\mathcal{L}$

$$\mathcal{L}G(x) \stackrel{\text{def}}{=} \mathcal{T}G(x(\sigma)) = \sum_{\eta \in \{+, -\}} G(F_\eta(x)) e^{-\beta x_\eta}, \quad (\text{A33.39})$$

where all dependencies on σ have disappeared — if we know the value of the state observables x , the action of $\mathcal{L}$ on G can be computed.

A dynamical system is formed out of the maps F_η . They are chosen so that one of the state variables is the interaction energy. One can consider the

two maps F_+ and F_- as the inverse branches of a hyperbolic map f , that is, $f^{-1}(x) = \{F_+(x), F_-(x)\}$. Studying the thermodynamics of the interaction ϕ is equivalent to studying the long term behavior of the orbits of the map f , achieving the transformation of the spin system into a dynamical system.

Unlike the original transfer operator, the $\mathcal{L}$ operator — acting in the space of observables that depend only on the state variables — is of trace-class (its trace is finite). The finite trace gives us a chance to relate the trace of $\mathcal{L}^n$ to the partition function of a system of size n . We can do better. As most properties of interest (thermodynamics, fall-off of correlations) are determined directly from its spectrum, we can study instead the zeros of the Fredholm determinant $\det(1 - z\mathcal{L})$ by the technique of cycle expansions developed for dynamical systems [2]. A cycle expansion consists of finding a power series expansion for the determinant by writing $\det(1 - z\mathcal{L}) = \exp(\text{tr} \ln(1 - z\mathcal{L}))$. The logarithm is expanded into a power series and one is left with terms of the form $\text{tr} \mathcal{L}^n$ to evaluate. For evaluating the trace, the $\mathcal{L}$ operator is equivalent to

$$\mathcal{L}G(x) = \int_{\mathbf{R}^D} dy \delta(y - f(x)) e^{-\beta y} G(y) \quad (\text{A33.40})$$

from which the trace can be computed:

$$\text{tr} \mathcal{L}^n = \sum_{x=f^{(on)}(x)} \frac{e^{-\beta H(x)}}{|\det(1 - \partial_x f^{(on)}(x))|} \quad (\text{A33.41})$$

with the sum running over all the fixed points of $f^{(on)}$ (all spin configurations of a given length). Here $f^{(on)}$ is f composed with itself n times, and $H(x)$ is the energy of the configuration associated with the point x . In practice the map f is never constructed and the energies are obtained directly from the spin configurations.

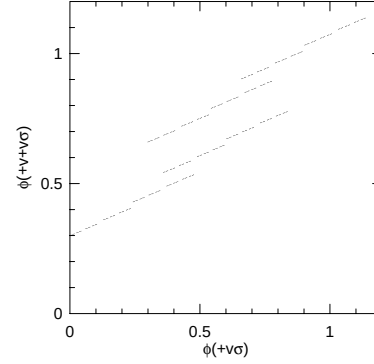
To compute the value of $\text{tr} \mathcal{L}^n$ we must compute the value of $\partial_x f^{(on)}$; this involves a functional derivative. To any degree of accuracy a number x in the range of possible interaction energies can be represented by a finite string of spins ϵ , such as $x = \phi(+, \epsilon_0, \epsilon_1, \dots, -, -, \dots)$. By choosing the sequence ϵ to have a large sequence of spins $-$, the number x can be made as small as needed, so in particular we can represent a small variation by $\phi(\eta)$. As $x_+(\epsilon) = \phi(+ \vee \epsilon)$, from the definition of a derivative we have:

$$\partial_x f(x) = \lim_{m \rightarrow \infty} \frac{\phi(\epsilon \vee \eta^{(m)}) - \phi(\epsilon)}{\phi(\eta^{(m)})}, \quad (\text{A33.42})$$

where $\eta^{(m)}$ is a sequence of spin strings that make $\phi(\eta^{(m)})$ smaller and smaller. By substituting the definition of ϕ in terms of its pairwise interaction $J(n) = n^s a^{n^\gamma}$ and taking the limit for the sequences $\eta^{(m)} = \{+, -, -, \dots, \eta_{m+1}, \eta_{m+2}, \dots\}$ one computes that the limit is a if $\gamma = 1$, 1 if $\gamma < 1$, and 0 if $\gamma > 1$. It does not depend on the positive value of s . When $\gamma < 1$ the resulting dynamical system is not hyperbolic and the construction for the operator $\mathcal{L}$ fails, so one cannot apply it to potentials such as $(1/2)^{\sqrt{n}}$. One may solve this problem by investigating the behavior of the formal dynamical system as $\gamma \rightarrow 0$.

The manipulations have up to now assumed that the map f is smooth. If the dimension D of the embedding space is too small, f may not be smooth.

Figure A33.3: The spin adding map F_+ for the potential $J(n) = \sum n^2 a^n$. The action of the map takes the value of the interaction energy between $+$ and the semi-infinite configuration $\{\sigma_1, \sigma_2, \sigma_3, \dots\}$ and returns the interaction energy between $+$ and the configuration $\{+, \sigma_1, \sigma_2, \sigma_3, \dots\}$.



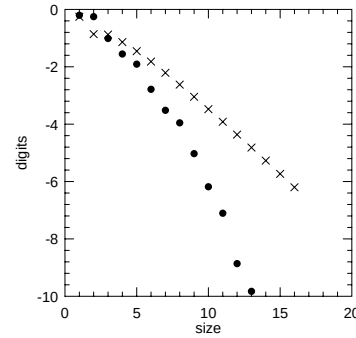
Determining under which conditions the embedding is smooth is a complicated question [18]. But in the case of spin systems with pairwise interactions it is possible to give a simple rule. If the interaction is of the form

$$\phi(\sigma) = \sum_{n \geq 1} \delta_{\sigma_0, \sigma_n} \sum_k p_k(n) a_k^{n^\gamma} \quad (\text{A33.43})$$

where p_k are polynomials and $|a_k| < 1$, then the state observables to use are $x_{s,k}(\sigma) = \sum \delta_{+, \sigma_n} n^s a_k^n$. For each k one uses $x_{0,k}, x_{1,k}, \dots$ up to the largest power in the polynomial p_k . An example is the interaction with $J_1(n) = n^2(3/10)^n$. It leads to a 3-dimensional system with variables $x_{0,0}, x_{1,0}$, and $x_{2,0}$. The action of the map F_+ for this interaction is illustrated figure A33.3. Plotted are the pairs $\{\phi(+ \vee \sigma), \phi(+ \vee + \vee \sigma)\}$. This can be seen as the strange attractor of a chaotic system for which the variables $x_{0,0}, x_{1,0}$, and $x_{2,0}$ provide a good (analytic) embedding.

The added smoothness and trace-class of the $\mathcal{L}$ operator translates into faster convergence towards the thermodynamic limit. As the reconstructed dynamics is analytic, the convergence towards the thermodynamic limit is faster than exponential [5, 16]. We will illustrate this with the polynomial-exponential interactions (A33.43) with $\gamma = 1$, as the convergence is certainly faster than exponential if $\gamma > 1$, and the case of a^n has been studied in terms of another Fredholm determinant by Gutzwiller [9]. The convergence is illustrated in figure A33.4 for the interaction $n^2(3/10)^n$. Plotted in the graph, to illustrate the transfer matrix convergence, are the number of decimal digits that remain unchanged as the range of the interaction is increased. Also in the graph are the number of decimal digits that remain unchanged as the largest power of $\text{tr } \mathcal{L}^n$ considered. The plot is effectively a logarithmic plot and straight lines indicate exponentially fast convergence. The curvature indicates that the convergence is faster than exponential. By fitting, one can verify that the free energy is converging to its limiting value as $\exp(-n^{(4/3)})$. Cvitanović [5] has estimated that the Fredholm determinant of a map on a D dimensional space should converge as $\exp(-n^{(1+1/D)})$, which is confirmed by these numerical simulations.

Figure A33.4: Number of digits for the Fredholm method (●) and the transfer function method (×). The size refers to the largest cycle considered in the Fredholm expansions, and the truncation length in the case of the transfer matrix.



Résumé

The geometrization of spin systems strengthens the connection between statistical mechanics and dynamical systems. It also further establishes the value of the Fredholm determinant of the $\mathcal{L}$ operator as a practical computational tool with applications to chaotic dynamics, spin systems, and semiclassical mechanics. The example above emphasizes the high accuracy that can be obtained: by computing the shortest 14 periodic orbits of period 5 or less it is possible to obtain three digit accuracy for the free energy. For the same accuracy with a transfer matrix one has to consider a 256×256 matrix. This makes the method of cycle expansions practical for analytic calculations.

Commentary

Remark A33.1. Presentation functions. The best place to read about Feigenbaum's work is in his review article published in *Los Alamos Science* (reproduced in several reprint collections and conference proceedings, such as ref. [4]). Feigenbaum's *J. Stat. Phys.* article [6] is the easiest place to learn about presentation functions.

Remark A33.2. Interactions are smooth In most computational schemes for thermodynamic quantities the translation invariance and the smoothness of the basic interaction are never used. In Monte Carlo schemes, aside from the periodic boundary conditions, the interaction can be arbitrary. In principle for each configuration it could be possible to have a different energy. Schemes such as the Swenson-Wang cluster flipping algorithm use the fact that interaction is local and are able to obtain dramatic speed-ups in the equilibration time for the dynamical Monte Carlo simulation. In the geometrization program for spin systems, the interactions are assumed translation invariant and smooth. The smoothness means that any interaction can be decomposed into a series of terms that depend only on the spin arrangement and the distance between spins:

$$\phi(\sigma_0, \sigma_1, \sigma_2, \dots) = J_0 \sigma_0 + \sum \delta(\sigma_0, \sigma_n) J_1(n) + \sum \delta(\sigma_0, \sigma_{n_1}, \sigma_{n_2}) J_2(n_1, n_2) + \dots$$

where the J_k are symmetric functions of their arguments and the δ are arbitrary discrete functions. This includes external constant fields (J_0), but it excludes site dependent fields such as a random external magnetic field.

References

- [1] R. Artuso, “Logarithmic strange sets”, *J. Phys. A* **21**, L923–L927 (1988).
- [2] R. Artuso, E. Aurell, and P. Cvitanović, “Recycling of strange sets: I. Cycle expansions”, *Nonlinearity* **3**, 325–359 (1990).
- [3] M. F. Barnsley, *Fractals Everywhere* (Dover, New York, 2012).
- [4] P. Cvitanović, *Universality in Chaos*, 2nd ed. (Adam Hilger, Bristol, 1989).
- [5] P. Cvitanović, “Periodic orbits as the skeleton of classical and quantum chaos”, *Physica D* **51**, 138–151 (1991).
- [6] M. J. Feigenbaum, “Presentation functions, fixed points, and a theory of scaling function dynamics”, *J. Stat. Phys.* **52**, 527–569 (1988).
- [7] M. E. Fisher, “The theory of condensation and the critical point”, *Physics* **3**, 255–283 (1967).
- [8] G. Gallavotti, “Funzioni zeta ed insiemi basilari”, *Accad. Lincei. Rend. Sc. Fis. Mat. e Nat* **61**, 309–317 (1976).
- [9] M. C. Gutzwiller, “The quantization of a classically ergodic system”, *Physica D* **5**, 183–207 (1982).
- [10] J. Hutchinson, “Fractals and self-similarity”, *Indiana Univ. Math. J.* **30**, 713–747 (1981).
- [11] H. A. Kramers and G. H. Wannier, “Statistics of the two-dimensional ferromagnet. Part I”, *Phys. Rev.* **60**, 252–262 (1941).
- [12] D. H. Mayer, *The Ruelle-Araki Transfer Operator in Classical Statistical Mechanics* (Springer, Berlin, 1980).
- [13] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, “Geometry from a time series”, *Phys. Rev. Lett.* **45**, 712–716 (1980).
- [14] D. Ruelle, “Statistical mechanics of a one-dimensional lattice gas”, *Commun. Math. Phys.* **9**, 267–288 (1968).
- [15] D. Ruelle, *Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics*, 2nd ed. (Cambridge Univ. Press, Cambridge UK, 2004).
- [16] H. H. Rugh, Time Evolution and Correlations in Chaotic Dynamical Systems, PhD thesis (Univ. of Copenhagen, Copenhagen, 1992).
- [17] F. Takens, “Detecting strange attractors in turbulence”, in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L.-S. Young (Springer, New York, 1981), pp. 366–381.
- [18] P. Walters, *An Introduction to Ergodic Theory* (Springer, New York, 1981).

A33.6 Exercises

- A33.1. **Not all Banach spaces are also Hilbert.** If we are given a norm $\|\cdot\|$ of a Banach space B , it may be possible to find an inner product $\langle \cdot, \cdot \rangle$ (so that B is also a Hilbert space H) such that for all vectors $f \in B$, we have

$$\|f\| = \langle f, f \rangle^{1/2}.$$

This is the norm induced by the scalar product. If we cannot find the inner product how do we know that we just are not being clever enough? By checking the parallelogram law for the norm. A Banach space can be made into a Hilbert space if and only if the norm satisfies the parallelogram law. The parallelogram law says that for any two vectors f and g the equality

$$\|f + g\|^2 + \|f - g\|^2 = 2\|f\|^2 + 2\|g\|^2,$$

must hold.

Consider the space of bounded observables with the norm given by $\|a\| = \sup_{\sigma \in \Omega^{\mathbb{N}}} |a(\sigma)|$. Show that there is no scalar product that will induce this norm.

- A33.2. **Automaton for a droplet.** Find the transition graph and the weights on the edges so that the energies of configurations for the droplet model are correctly generated. For any string starting in zero and ending in zero your diagram should yield a configuration the weight $e^{H(\sigma)}$, with H computed along the lines of (A33.13) and (A33.18).

Hint: the transition graph is infinite.

- A33.3. **Spectral determinant for a^n interactions.** Compute the spectral determinant for 1-dimensional Ising model with the interaction

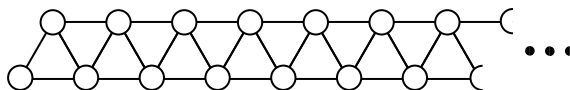
$$\phi(\sigma) = \sum_{k>0} a^k \delta(\sigma_0, \sigma_k).$$

Take a as a number smaller than $1/2$.

- What is the dynamical system this generates? That is, find F_+ and F_- as used in (A33.39).
- Show that

$$\frac{d}{dx} F_{\{+ \text{ or } -\}} = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$$

- A33.4. **Ising model on a thin strip.** Compute the transfer matrix for the Ising model defined on the graph



Assume that whenever there is a bond connecting two sites, there is a contribution $J\delta(\sigma_i, \sigma_j)$ to the energy.

- A33.5. **Infinite symbolic dynamics.** Let σ be a function that returns zero or one for every infinite binary string: $\sigma : \{0, 1\}^{\mathbb{N}} \rightarrow \{0, 1\}$. Its value is represented by $\sigma(\epsilon_1, \epsilon_2, \dots)$ where the ϵ_i are

either 0 or 1. We will now define an operator $\mathcal{T}$ that acts on observables on the space of binary strings. A function a is an observable if it has bounded variation, that is, if

$$\|a\| = \sup_{\{\epsilon_i\}} |a(\epsilon_1, \epsilon_2, \dots)| < \infty.$$

For these functions

$$\begin{aligned} \mathcal{T}a(\epsilon_1, \epsilon_2, \dots) &= a(0, \epsilon_1, \epsilon_2, \dots)\sigma(0, \epsilon_1, \epsilon_2, \dots) \\ &\quad + a(1, \epsilon_1, \epsilon_2, \dots)\sigma(1, \epsilon_1, \epsilon_2, \dots). \end{aligned} \quad (\text{A33.44})$$

The function σ is assumed such that any of $\mathcal{T}$'s “matrix representations” in (a) have the Markov property (the matrix, if read as an adjacency graph, corresponds to a graph where one can go from any node to any other node).

- (a) (easy) Consider a finite version T_n of the operator $\mathcal{T}$:

$$\begin{aligned} T_n a(\epsilon_1, \epsilon_2, \dots, \epsilon_n) &= \\ &\quad a(0, \epsilon_1, \epsilon_2, \dots, \epsilon_{n-1})\sigma(0, \epsilon_1, \epsilon_2, \dots, \epsilon_{n-1}) + \\ &\quad a(1, \epsilon_1, \epsilon_2, \dots, \epsilon_{n-1})\sigma(1, \epsilon_1, \epsilon_2, \dots, \epsilon_{n-1}). \end{aligned}$$

Show that T_n is a $2^n \times 2^n$ matrix. Show that its trace is bounded by a number independent of n .

- (b) (medium) With the operator norm induced by the function norm, show that $\mathcal{T}$ is a bounded operator.
- (c) (hard) Show that $\mathcal{T}$ is not trace-class. (Hint: check if $\mathcal{T}$ is compact).

Classes of operators are nested; trace-class $\leq$ compact $\leq$ bounded.

Appendix A37

WKB quantization

(H. Liang and P. Cvitanović)

THE SEMICLASSICAL FIELD THEORY, derived below as a WKB approximation to a quantum path integral, and Gutzwiller's semiclassical quantization, both have support on the set of deterministic solutions of system's defining equations, solutions which we refer to as *periodic states*.

chapter 37

To keep things simple (you know what the determinant of a matrix is, but what's a *functional* determinant?), we introduce the theory by discretizing it over a hypercubic integer lattice (A24.11), and consider first a finite-dimensional state space (A24.18) over primitive cell $\mathbb{A}$ (A24.17) of finite volume (A24.19).

section A24.1.1

In the path integral formulation of quantum field theory, a field configuration Φ (A24.13) over primitive cell $\mathbb{A}$ occurs with probability amplitude density

$$p_{\mathbb{A}}[\Phi] = \frac{1}{Z_{\mathbb{A}}} e^{\frac{i}{\hbar} S[\Phi]}, \quad Z_{\mathbb{A}} = Z_{\mathbb{A}}[0], \quad (\text{A37.1})$$

where $S[\Phi]$ is the action of the field configuration Φ . The $V_{\mathbb{A}}$ -dimensional primitive cell partition sum is the integral over all field configurations over primitive cell $\mathbb{A}$

$$Z_{\mathbb{A}}[\mathbf{J}] = \int d\Phi_{\mathbb{A}} e^{\frac{i}{\hbar} (S[\Phi] + \Phi \cdot \mathbf{J})}, \quad d\Phi_{\mathbb{A}} = \prod_{z \in \mathbb{A}} \frac{d\phi_z}{\sqrt{2\pi}}. \quad (\text{A37.2})$$

Here the 'sources' $\mathbf{J} = \{J_z\}$ are added to the action to facilitate the evaluation of expectation values of field moments (n -point Green's functions) by applications of d/dJ_z to the partition sum (A37.2):

$$\langle \phi_i, \phi_j, \dots, \phi_n \rangle_{\mathbb{A}} = \int d\Phi_{\mathbb{A}} \phi_i \phi_j \dots \phi_n p_{\mathbb{A}}[\Phi], \quad (\text{A37.3})$$

A *semiclassical* (or *WKB*) approximation to the partition sum is obtained by the method of stationary phase. We illustrate this by a 0-dimensional lattice field theory.

A37.1 Semiclassical field theory, a single lattice site

Consider a Laplace integral of form

section 37.2

$$\langle a \rangle_0 = \int \frac{d\phi}{\sqrt{2\pi}} a(\phi) e^{\frac{i}{\hbar} S(\phi)}, \quad (\text{A37.4})$$

with a real-valued positive parameter $\hbar$, a real-valued function $S(\phi)$, and an observable $a(\phi)$. Laplace estimate of this integral is obtained by determining its extremal point ϕ_c , given by the stationary phase condition

$$\frac{d}{d\phi} S(\phi_c) = 0, \quad (\text{A37.5})$$

and approximating the action to second order,

$$S(\phi) = S(\phi_c) + \frac{1}{2} S''(\phi_c) (\phi - \phi_c)^2 + \dots$$

The contribution of the quadratic term is given by the Fresnel integral (37.16)

exercise 37.2

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\phi e^{-\frac{\phi^2}{2ib}} = \sqrt{ib} = |b|^{1/2} e^{i\frac{\pi}{4} \frac{b}{|b|}}, \quad (\text{A37.6})$$

with $b = \hbar/S''(\phi_c)$ and the phase depending on the sign of $S''(\phi_c)$, so for a lattice with a single site the *semiclassical* approximation to the partition sum formula (A37.4) for the expectation value is

$$\langle a \rangle_0 = \int \frac{d\phi}{\sqrt{2\pi}} a(\phi) e^{\frac{i}{\hbar} S(\phi)} \approx a(\phi_c) \frac{e^{\frac{i}{\hbar} S(\phi_c) \pm i\frac{\pi}{4}}}{|S''(\phi_c)/\hbar|^{1/2}}, \quad (\text{A37.7})$$

with $\pm$ for positive/negative sign of $S''(\phi_c)$.

A37.2 Semiclassical lattice field theory

The semiclassical approximation to the lattice field theory partition sum (A37.2) is a $V_{\mathbb{A}}$ -dimensional generalization of the above Laplace-Fresnel integral. The stationary phase condition (A37.5)

$$F[\Phi_c]_z = \frac{\delta S[\Phi_c]}{\delta \phi_z} = 0 \quad (\text{A37.8})$$

is Euler–Lagrange equation, whose global *deterministic* solutions Φ_c satisfy this local extremal condition on every lattice site z ; in system's state space (A24.18) Φ_c is a stationary *point* of the action $S[\Phi]$.

In the *WKB approximation*, the action near the point Φ_c is expanded to quadratic order,

$$S[\Phi] \approx S[\Phi_c] + \frac{1}{2} (\Phi - \Phi_c)^\top \mathcal{J}_c (\Phi - \Phi_c), \quad (\text{A37.9})$$

where we refer to the matrix of second derivatives

$$(\mathcal{J}_c)_{z'z} = \left. \frac{\delta^2 S[\Phi]}{\delta \phi_{z'} \delta \phi_z} \right|_{\Phi=\Phi_c} \quad (\text{A37.10})$$

as the *orbit Jacobian matrix*. The Fresnel integral (A37.6) is now a multidimensional integral over $V_{\mathbb{A}}$ lattice sites state-space neighborhood $\mathcal{M}_c$ of a deterministic solution Φ_c approximated by a Gaussian

$$\begin{aligned} \int d\Phi_{\mathbb{A}} e^{\frac{i}{\hbar} \Phi^\top \mathcal{J}_c \Phi} &= \frac{1}{|\text{Det}(\mathcal{J}_c/\hbar)|^{1/2}} e^{im_c}, \\ d\Phi_{\mathbb{A}} &= \prod_{z \in \mathbb{A}} \frac{d\phi_z}{\sqrt{2\pi}}, \end{aligned} \quad (\text{A37.11})$$

where the Maslov index m_c is a sum of phases (A37.6), with signs determined by the signs of eigenvalues of $\mathcal{J}_c$.

Our semiclassical d -dimensional spatiotemporal quantum *field* theory is a generalization of Gutzwiller [4] semiclassical approximation to quantum *mechanics* (temporal quantum evolution of a classically low-dimensional mechanical system, no infinite spatial directions). It assigns a *quantum* probability amplitude to a *deterministic* solution Φ_c , [2, 5–7]

$$p_c(\Phi) \approx \frac{1}{Z_{\mathbb{A}}} \frac{e^{\frac{i}{\hbar} S[\Phi_c] + im_c}}{|\text{Det}(\mathcal{J}_c/\hbar)|^{1/2}}, \quad Z_{\mathbb{A}} = Z_{\mathbb{A}}[0], \quad (\text{A37.12})$$

with the partition sum (A37.2) having support on the set of *deterministic multiperiodic solutions* Φ_c over primitive cell $\mathbb{A}$,

$$Z_{\mathbb{A}}[\mathbf{J}] \approx \sum_c \frac{e^{\frac{i}{\hbar} S[\Phi_c] + im_c + i\Phi_c \cdot \mathbf{J}}}{|\text{Det}(\mathcal{J}_c/\hbar)|^{1/2}}. \quad (\text{A37.13})$$

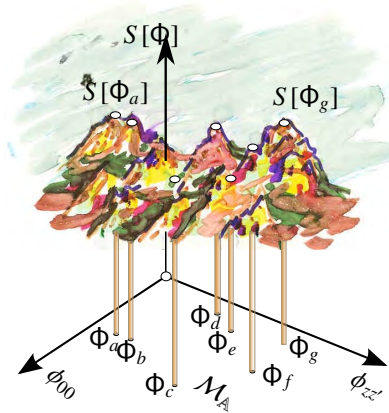
The actual calculations are always carried out over the infinite lattice, more precisely (not standard in solid state physics, but necessary to describe a chaotic field theory) over the set of all periodic lattice field configurations (A24.16) over all Bravais lattices $\mathcal{L}_{\mathbb{A}}$ (A24.17), or, in field theory, as the ‘sum over geometries’.

section A24.1.2

Résumé

The backbone of semiclassical *quantum* field theory is the set of *deterministic* solutions of system’s defining equations (A37.8), with the leading exponential contribution given by action evaluated on the deterministic solution, while the next-to-leading prefactor is the determinant of the operator describing quantum fluctuations about the deterministic solution.

Figure A37.1: A bird’s eye view of the $S[\Phi]$ action landscape over the primitive cell state space $\mathcal{M}_A$ (A24.18). White ellipses indicate the stationary points (A37.8), i.e., the set of all deterministic solutions $\{\Phi_a, \Phi_b, \Phi_c, \dots, \Phi_g\}$. They form the skeleton on which the partition sums of quantum chaos are evaluated, with the common deterministic backbone, but different weights. For a quantum theory, the semiclassical partition sum (A37.13) is an approximation, with quantum probability amplitude phases given by deterministic solutions’ actions, and stability weights given by square roots of the deterministic ones.



Commentary

Remark A37.1. Variations on the WKB theme. We could have equally well derived the Onsager-Machlup-Freidlin-Wentzell [3] weak noise saddle-point approximation, or Castro *et al.* [1] the semiclassical limit of the pure Einstein gravity partition function, and arrived to the same conclusion: in such limits partition sums have support on the set of deterministic periodic solutions.

Question A37.1. Alice, baffled

Q Why does this derivation of semiclassical field theory (PDEs!) seems so easy, when the quantum-mechanical WKB of chapter 37 is such a mess?

A Wisdom of Lagrange, Dirac, Feynman: we work in Lagrangian formulation of spatio-temporal field theory. The Hamiltonian formulation (37.2) would be a pain. You would do us, the authors as well as the readers, a great favor if you re-derived the results of chapter 37 in a Lagrangian setting.

References

- [1] A. Castro, M. R. Gaberdiel, T. Hartman, A. Maloney, and R. Volpato, “Gravity dual of the Ising model”, *Phys. Rev. D* **85**, 024032 (2012).
- [2] Y. Colin de Verdière, “Spectrum of the Laplace operator and periodic geodesics: thirty years after”, *Ann. Inst. Fourier* **57**, 2429–2463 (2007).
- [3] M. I. Freidlin and A. D. Wentzel, *Random Perturbations of Dynamical Systems* (Springer, Berlin, 1998).
- [4] M. C. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer, New York, 1990).
- [5] S. Levit and U. Smilansky, “A new approach to Gaussian path integrals and the evaluation of the semiclassical propagator”, *Ann. Phys.* **103**, 198–207 (1977).
- [6] S. Levit and U. Smilansky, “A theorem on infinite products of eigenvalues of Sturm-Liouville type operators”, *Proc. Amer. Math. Soc.* **65**, 299–299 (1977).

- [7] J. H. Van Vleck, “[The correspondence principle in the statistical interpretation of quantum mechanics](#)”, *Proc. Natl. Acad. Sci.* **14**, 178–188 (1928).

Appendix A39

Semiclassical quantization, with corrections

In fact I'm Vattay Gábor, just these Indo-Europeans mix up the right order.

(G. Vattay)

THE GUTZWILLER TRACE FORMULA is only a good approximation to the quantum mechanics when $\hbar$ is small. Can we improve the trace formula by adding quantum corrections to the semiclassical terms? A similar question can be posed when the deterministic dynamics is disturbed by some way Gaussian white noise with strength D . The deterministic dynamics then can be considered as the weak noise limit $D \rightarrow 0$. The effect of the noise can be taken into account by adding noise corrections to the deterministic trace formula. A formal analogy exists between the noise and the quantum problem. This analogy allows us to treat the noise and quantum corrections together.



A39.1 Periodic orbits as integrable systems

From now on, we use the language of quantum mechanics, since it is more convenient to visualize the results there. Where it is necessary we will discuss the difference between noise and quantum cases.

First, we introduce periodic orbits from an unusual point of view, which can convince you, that chaotic and integrable systems are in fact not as different from each other, than we might think. If we start orbits in the neighborhood of a periodic orbit and look at the picture on the Poincaré section we can see a regular picture. For stable periodic orbits the points form small ellipses around the center and for unstable orbits they form hyperbolas (see figure [A39.1](#)).

Figure A39.1: Poincaré section close to a stable and an unstable periodic orbit.

The motion close to a periodic orbit is regular in both cases. This is due to the fact, that we can linearize the Hamiltonian close to an orbit, and linear systems are always integrable. The linearized Hamilton's equations close to the periodic orbit $(q_p(t) + q, p_p(t) + p)$ are of the form

$$\begin{aligned}\dot{q} &= +\partial_{pq}^2 H(q_p(t), p_p(t)) q + \partial_{pp}^2 H(q_p(t), p_p(t)) p \\ \dot{p} &= -\partial_{qq}^2 H(q_p(t), p_p(t)) q - \partial_{qp}^2 H(q_p(t), p_p(t)) p,\end{aligned}$$

where the new coordinates q and p are relative to a periodic orbit. This linearized equation can be regarded as a D -degrees of freedom oscillator with time periodic frequencies. These equations are representing the equation of motion in a redundant way since more than one combination of q , p and t determines the same point of the phase space. This can be cured by an extra restriction on the variables, a constraint the variables should fulfill. This constraint can be derived from the time independence or stationarity of the full Hamiltonian

$$\partial_t H(q_p(t) + q, p_p(t) + p) = 0. \quad (\text{A39.1})$$

Using the linearized form of this constraint we can eliminate one of the linearized equations. It is very useful, although technically difficult, to do one more transformation and to introduce a coordinate, which is parallel with the Hamiltonian flow ($x_{||}$) and others which are orthogonal. In the orthogonal directions we again get linear equations. These equations with $x_{||}$ dependent rescaling can be transformed into normal coordinates, so that we get tiny oscillators in the new coordinates with constant frequencies. This result has first been derived by Poincaré for equilibrium points and later it was extended for periodic orbits by V.I. Arnol'd and co-workers. In the new coordinates, the Hamiltonian reads as

$$H_0(x_{||}, p_{||}, x_n, p_n) = \frac{1}{2} p_{||}^2 + U(x_{||}) + \sum_{n=1}^{D-1} \frac{1}{2} (p_n^2 \pm \omega_n^2 x_n^2), \quad (\text{A39.2})$$

which is the general form of the Hamiltonian in the neighborhood of a periodic orbit. The $\pm$ sign denotes, that for stable modes the oscillator potential is positive while for an unstable mode it is negative. For the unstable modes, ω is the Lyapunov exponent of the orbit

$$\omega_n = \ln \Lambda_{p,n} / T_p, \quad (\text{A39.3})$$

where $\Lambda_{p,n}$ is the expanding eigenvalue of the Jacobi matrix. For the stable directions the eigenvalues of the Jacobi matrix are connected with ω as

$$\Lambda_{p,n} = e^{-i\omega_n T_p}. \quad (\text{A39.4})$$

The Hamiltonian close to the periodic orbit is integrable and can be quantized by the Bohr-Sommerfeld rules. The result of the Bohr-Sommerfeld quantization for

the oscillators gives the energy spectra

$$\begin{aligned} E_n &= \hbar\omega_n \left(j_n + \frac{1}{2} \right) && \text{for stable modes} \\ E_n &= -i\hbar\omega_n \left(j_n + \frac{1}{2} \right) && \text{for unstable modes,} \end{aligned} \quad (\text{A39.5})$$

where $j_n = 0, 1, \dots$. It is convenient to introduce the index $s_n = 1$ for stable and $s_n = -i$ for unstable directions. The parallel mode can be quantized implicitly through the deterministic action function of the mode:

$$\frac{1}{2\pi} \oint p_{\parallel} dx_{\parallel} = \frac{1}{2\pi} S_{\parallel}(E_m) = \hbar \left(m + \frac{m_p \pi}{2} \right), \quad (\text{A39.6})$$

where m_p is the topological index of the motion in the parallel direction. The latter condition can be rewritten by a very useful trick into the equivalent form (39.18)

$$1 - e^{iS_{\parallel}(E_m)/\hbar - im_p \pi/2} = 0. \quad (\text{A39.7})$$

The eigenenergies of a semiclassically quantized periodic orbit are all the possible energies

$$E = E_m + \sum_{n=1}^{D-1} E_n. \quad (\text{A39.8})$$

This relation allows us to change in (A39.7) E_m with the full energy minus the oscillator energies $E_m = E - \sum_n E_n$. All the possible eigenenergies of the periodic orbit then are the zeroes of the expression

$$\Delta_p(E) = \prod_{j_1, \dots, j_{D-1}} (1 - e^{iS_{\parallel}(E - \sum_n \hbar s_n \omega_n (j_n + 1/2))/\hbar - im_p \pi/2}). \quad (\text{A39.9})$$

If we Taylor expand the action around E to first order

$$S_{\parallel}(E + \epsilon) \approx S_{\parallel}(E) + T(E)\epsilon, \quad (\text{A39.10})$$

where $T(E)$ is the period of the orbit, and use the relations of ω and the eigenvalues of the Jacobi matrix, we get the expression of the Selberg product

$$\Delta_p(E) = \prod_{j_1, \dots, j_{D-1}} \left(1 - \frac{e^{iS_p(E)/\hbar - im_p \pi/2}}{\prod_n \Lambda_{p,n}^{(1/2 + j_n)}} \right). \quad (\text{A39.11})$$

If we use the right convention for the square root we get exactly the D -degrees of freedom expression of the Selberg product formula we derived from the Gutzwiller trace formula in (39.19). Just here we derived it in a different way! The function $\Delta_p(E)$ is the semiclassical zeta function for one prime orbit.

Now, if we have many prime orbits and we would like to construct a function which is zero, whenever the energy coincides with the Bohr-Sommerfeld quantized energy of one of the periodic orbits, we have to take the product of these determinants:

$$\Delta(E) = \prod_p \Delta_p(E). \quad (\text{A39.12})$$

The miracle of the semiclassical zeta function is, that if we take infinitely many periodic orbits, the infinite product will have zeroes not at these energies, but close to the eigenenergies of the whole system!

So we have learned that both stable and unstable orbits are integrable systems, and can be individually quantized semiclassically by the old Bohr-Sommerfeld rules. We have thus almost completed the program of Sommerfeld to quantize general systems with the method of Bohr. *A remark: In addition to the Bohr-Sommerfeld rules, we used the unjustified approximation (A39.10). Sommerfeld would never do this! At that point we loose some important precision compared to the Bohr-Sommerfeld rules and we get somewhat worse results than a semi-classical formula is able to do. We will come back to this point in sect. A39.4, when we discuss the quantum corrections.* To complete the program of full scale Bohr-Sommerfeld quantization of chaotic systems we have to go beyond the linear approximation around the periodic orbit.

The Hamiltonian close to a periodic orbit in the parallel and normal coordinates can be written as the ‘harmonic’ plus ‘anaharmonic’ perturbation

$$H(x_{||}, p_{||}, x_n, p_n) = H_0(x_{||}, p_{||}, x_n, p_n) + H_A(x_{||}, x_n, p_n), \quad (\text{A39.13})$$

where the anaharmonic part can be written as a sum of homogeneous polynomials of x_n and p_n with $x_{||}$ dependent coefficients,

remark A39.1

$$\begin{aligned} H_A(x_{||}, x_n, p_n) &= \sum_{k=3} H^k(x_{||}, x_n, p_n) \\ H^k(x_{||}, x_n, p_n) &= \sum_{\ell_n + m_n = k} H_{\ell_n, m_n}^k(x_{||}) x_n^{\ell_n} p_n^{m_n}. \end{aligned} \quad (\text{A39.14})$$

This deterministic Hamiltonian is hopeless from Sommerfeld’s point of view, since it is non integrable. However, Birkhoff in 1927 introduced the concept of normal form, which gives successive integrable approximations to a non-integrable problem.³

A39.2 The Birkhoff normal form

Birkhoff studied the canonical perturbation theory close to an equilibrium point of a Hamiltonian. Equilibrium point is where the potential has a minimum $\nabla U = 0$ and small perturbations lead to oscillatory motion. We can linearize the problem and by introducing normal coordinates x_n and conjugate momenta p_n the quadratic part of the Hamiltonian will be a set of oscillators

$$H_0(x_n, p_n) = \sum_{n=1}^D \frac{1}{2} (p_n^2 + \omega_n^2 x_n^2). \quad (\text{A39.15})$$

³It is a pity that in 1926 Schrödinger introduced wave mechanics, thus blocking the development of Sommerfeld’s concept.

The full Hamiltonian can be rewritten with the new coordinates

$$H(x_n, p_n) = H_0(x_n, p_n) + H_A(x_n, p_n), \quad (\text{A39.16})$$

where H_A is the anaharmonic part of the potential in the new coordinates. The anaharmonic part can be written as a series of homogeneous polynomials

$$\begin{aligned} H_A(x_n, p_n) &= \sum_{j=3}^{\infty} H^j(x_n, p_n) \\ H^j(x_n, p_n) &= \sum_{|\ell|+|m|=j} h_{\ell m}^j x^\ell p^m, \end{aligned}$$

where $h_{\ell m}^j$ are real constants and we used the multi-indices $\ell := (\ell_1, \dots, \ell_D)$ with definitions

$$|\ell| = \sum \ell_n, x^\ell := x_1^{\ell_1} x_2^{\ell_2} \dots x_D^{\ell_D}.$$

Birkhoff showed, that that by successive canonical transformations one can introduce new momenta and coordinates such, that in the new coordinates the anaharmonic part of the Hamiltonian up to any given n polynomial will depend only on the variable combination

$$\tau_n = \frac{1}{2}(p_n^2 + \omega_n^2 x_n^2), \quad (\text{A39.17})$$

where x_n and p_n are the new coordinates and momenta, but ω_n is the original frequency. This is called the Birkhoff normal form of degree N :

$$H(x_n, p_n) = \sum_{j=2}^N H^j(\tau_1, \dots, \tau_D), \quad (\text{A39.18})$$

where H^j are homogeneous degree j polynomials of τ 's. This is an integrable Hamiltonian, the non-integrability is pushed into the remainder, which consists of polynomials of degree higher than N . We run into trouble only when the oscillator frequencies are commensurate, i.e., when it is possible to find a set of integers m_n such that the linear combination

$$\sum_{n=1}^D \omega_n m_n,$$

vanishes. This extra problem has been solved by Gustavson in 1966 and we call the object Birkhoff-Gustavson normal form. The procedure of the successive canonical transformations can be computerized and can be carried out up to high orders (~ 20).

Of course, we pay a price for forcing the system to be integrable up to degree N . For a non-integrable system the high order terms behave quite wildly and the series is not convergent. Therefore we have to use this tool carefully. Now, we learned how to approximate a non-integrable system with a sequence of integrable systems and we can go back and carry out the Bohr-Sommerfeld quantization.

A39.3 Bohr-Sommerfeld quantization of periodic orbits

There is some difference between equilibrium points and periodic orbits. The Hamiltonian (A39.2) is not a sum of oscillators. One can transform the parallel part, describing circulation along the orbit, into an oscillator Hamiltonian, but this would make the problem extremely difficult. Therefore, we carry out the canonical transformations dictated by the Birkhoff procedure only in the orthogonal directions. The $x_{||}$ coordinate plays the role of a parameter. After the transformation up to order N the Hamiltonian (A39.14) is

$$H(x_{||}, p_{||}, \tau_1, \dots, \tau_{D-1}) = H_0(x_{||}, p_{||}, \tau_1, \dots, \tau_{D-1}) + \sum_{j=2}^N U^j(x_{||}, \tau_1, \dots, \tau_{D-1}), \quad (\text{A39.19})$$

where U^j is a j th order homogeneous polynomial of τ 's with $x_{||}$ dependent coefficients. The orthogonal part can be Bohr-Sommerfeld quantized by quantizing the individual oscillators, replacing τ 's as we did in (A39.5). This leads to a 1-dimensional effective potential indexed by $j_1, \dots, j_{D-1}$

$$\begin{aligned} H(x_{||}, p_{||}, j_1, \dots, j_{D-1}) &= \frac{1}{2} p_{||}^2 + U(x_{||}) + \sum_{n=1}^{D-1} \hbar s_n \omega_n (j_n + 1/2) \\ &+ \sum_{k=2}^N U^k(x_{||}, \hbar s_1 \omega_1 (j_1 + 1/2), \hbar s_2 \omega_2 (j_2 + 1/2), \dots, \hbar s_{D-1} \omega_{D-1} (j_{D-1} + 1/2)), \end{aligned} \quad (\text{A39.20})$$

where j_n can be any non-negative integer. The term with index k is proportional with $\hbar^k$ due to the homogeneity of the polynomials.

The parallel mode now can be Bohr-Sommerfeld quantized for any given set of j 's

$$\begin{aligned} S_p(E, j_1, \dots, j_{D-1}) &= \oint dx_{||} p_{||} \\ &= \oint dx_{||} \sqrt{E - \sum_{n=1}^{D-1} \hbar s_n \omega_n (j_n + 1/2) - U(x_{||}, j_1, \dots, j_{D-1})} \\ &= 2\pi\hbar(m + m_p/2), \end{aligned}$$

where U contains all the $x_{||}$ dependent terms of the Hamiltonian. The spectral determinant becomes

$$\Delta_p(E) = \prod_{j_1, \dots, j_{D-1}} (1 - e^{iS_p(E, j_1, \dots, j_{D-1})/\hbar - m_p\pi/2}).$$

This expression completes the Sommerfeld method and tells us how to quantize chaotic or general Hamiltonian systems. Unfortunately, Schrödinger's wave mechanics postponed this nice formula until our book.

The formula has been derived with the help of the semiclassical Bohr-Sommerfeld quantization rule and the deterministic normal form theory. Indeed, if we expand S_p in the exponent in the powers of $\hbar$

$$S_p = \sum_{k=0}^N \hbar^k S_k,$$

we get more than just a constant and a linear term. This formula already gives us corrections to the semiclassical zeta function in all powers of $\hbar$. There is a very attractive feature of this semiclassical expansion. $\hbar$ in S_p shows up only in the combination $\hbar s_n \omega_n(j_n + 1/2)$. A term proportional with $\hbar^k$ can only be a homogeneous expression of the oscillator energies $s_n \omega_n(j_n + 1/2)$. For example in two dimensions there is only one possibility of the functional form of the order k term

$$S_k = c_k(E) \cdot \omega_n^k(j + 1/2)^k,$$

where $c_k(E)$ is the only function to be determined.

The corrections derived so far are *doubly* semiclassical, since they give semiclassical corrections to the semiclassical approximation. What can quantum mechanics add to this? As we have stressed in the previous section, the exact quantum mechanics is not invariant under canonical transformations. In other context, this phenomenon is called the operator ordering problem. Since the operators $\hat{x}$ and $\hat{p}$ do not commute, we run into problems, when we would like to write down operators for deterministic quantities like $x^2 p^2$. On the deterministic level the four possible orderings $xpxp$, $ppxx$, $pxpx$ and $xxpp$ are equivalent, but they are different in the quantum case. The expression for the energy (A39.20) is not exact. We have to go back to the level of the Schrödinger equation in order to obtain the exact expression.

A39.4 Quantum calculation of $\hbar$ corrections

The Gutzwiller trace formula has originally been derived from the saddle point approximation of the Feynman path integral form of the propagator. The exact trace is a path-sum for all closed paths of the system

$$\text{Tr } G(x, x', t) = \int dx G(x, x, t) = \int \mathcal{D}x e^{iS(x, t)/\hbar}, \quad (\text{A39.21})$$

where $\int \mathcal{D}x$ denotes the discretization and summation for all paths of time length t in the limit of the infinite refinement and $S(x, t)$ is the deterministic action calculated along the path. The trace in the saddle point calculation is a sum for deterministic periodic orbits and zero length orbits, since these are the extrema of the action $\delta S(x, t) = 0$ for closed paths:

$$\text{Tr } G(x, x', t) = g_0(t) + \sum_{p \in PO} \int \mathcal{D}\xi_p e^{iS(\xi_p + x_p(t), t)/\hbar}, \quad (\text{A39.22})$$

where $g_0(t)$ is the zero length orbit contribution. We introduced the new coordinate ξ_p with respect to the periodic orbit $x_p(t)$, $x = \xi_p + x_p(t)$. Now, each path sum $\int \mathcal{D}\xi_p$ is computed in the vicinity of periodic orbits. Since the saddle points are taken in the configuration space, only spatially distinct periodic orbits, the so called prime periodic orbits, appear in the summation. So far nothing new has been invented. If we continue the standard textbook calculation scheme, we have to Taylor expand the action in ξ_p and keep the quadratic term in the exponent while treating the higher order terms as corrections. Then we can compute the path integrals with the help of Gaussian integrals. The key point here is that we do not compute the path sum directly. We use the correspondence between path integrals and partial differential equations. This idea comes from Maslov, and a good summary is given in ref. [2]. We search for that Schrödinger equation, which leads to the path sum

$$\int \mathcal{D}\xi_p e^{iS(\xi_p + x_p(t), t)/\hbar}, \quad (\text{A39.23})$$

where the action around the periodic orbit is in a multi-dimensional Taylor expanded form:

$$S(x, t) = \sum_{\mathbf{n}} s_{\mathbf{n}}(t) (x - x_p(t))^{\mathbf{n}} / \mathbf{n}! \quad (\text{A39.24})$$

The symbol $\mathbf{n} = (n_1, n_2, \dots, n_D)$ denotes the multi index for D -degrees of freedom, $\mathbf{n}! = \prod_{i=1}^D n_i!$ the multi factorial and $(x - x_p(t))^{\mathbf{n}} = \prod_{i=1}^D (x_i - x_{p,i}(t))^{n_i}$, respectively. The expansion coefficients of the action can be determined from the Hamilton-Jacobi equation

$$\partial_t S + \frac{1}{2}(\nabla S)^2 + U = 0, \quad (\text{A39.25})$$

in which the potential is expanded in a multidimensional Taylor series around the orbit

$$U(x) = \sum_{\mathbf{n}} u_{\mathbf{n}}(t) (x - x_p(t))^{\mathbf{n}} / \mathbf{n}!. \quad (\text{A39.26})$$

The Schrödinger equation

$$i\hbar \partial_t \psi = \hat{H}\psi = -\frac{\hbar^2}{2} \Delta \psi + U\psi, \quad (\text{A39.27})$$

with this potential also can be expanded around the periodic orbit. Using the WKB ansatz

$$\psi = \varphi e^{iS/\hbar}, \quad (\text{A39.28})$$

we can construct a Schrödinger equation corresponding to a given order of the Taylor expansion of the deterministic action. The Schrödinger equation induces the Hamilton-Jacobi equation (A39.25) for the phase and the transport equation of Maslov and Fjedorik [15] for the amplitude:

$$\partial_t \varphi + \nabla \varphi \nabla S + \frac{1}{2} \varphi \Delta S - \frac{i\hbar}{2} \Delta \varphi = 0. \quad (\text{A39.29})$$

This partial differential equation, solved in the neighborhood of a periodic orbit with the expanded action (A39.24), belongs to the local path-sum (A39.23).

If we know the Green's function $G_p(\xi, \xi', t)$ corresponding to the local equation (A39.29), then the local path sum can be converted back into a trace:

$$\int \mathcal{D}\xi_p e^{i/\hbar \sum_n S_n(x_p(t), t) \xi_p^n / n!} = \text{Tr } G_p(\xi, \xi', t). \quad (\text{A39.30})$$

The saddle point expansion of the trace in terms of local traces then becomes

$$\text{Tr } G(x, x', t) = \text{Tr } G_W(x, x', t) + \sum_p \text{Tr } G_p(\xi, \xi', t), \quad (\text{A39.31})$$

where $G_W(x, x', t)$ denotes formally the Green's function expanded around zero length orbits, known as the Weyl term [3–5]. Each Green's function can be Fourier-Laplace transformed independently and by definition we get in the energy domain

$$\text{Tr } G(x, x', E) = g_0(E) + \sum_p \text{Tr } G_p(\xi, \xi', E). \quad (\text{A39.32})$$

Note that we do not need here to take further saddle points in time, since we are dealing with exact time and energy domain Green's functions.

The spectral determinant is a function which has zeroes at the eigenenergies E_n of the Hamilton operator $\hat{H}$. Formally it is

$$\Delta(E) = \det(E - \hat{H}) = \prod_n (E - E_n).$$

The logarithmic derivative of the spectral determinant is the trace of the energy domain Green's function:

$$\text{Tr } G(x, x', E) = \sum_n \frac{1}{E - E_n} = \frac{d}{dE} \log \Delta(E). \quad (\text{A39.33})$$

We can define the spectral determinant $\Delta_p(E)$ also for the local operators, writing

$$\text{Tr } G_p(\xi, \xi', E) = \frac{d}{dE} \log \Delta_p(E). \quad (\text{A39.34})$$

Using (A39.32) we can express the full spectral determinant as a product for the sub-determinants

$$\Delta(E) = e^{W(E)} \prod_p \Delta_p(E),$$

where $W(E) = \int^E g_0(E') dE'$ is the term coming from the Weyl expansion.

The construction of the local spectral determinants can be done easily. We have to consider the stationary eigenvalue problem of the local Schrödinger problem and keep in mind, that we are in a coordinate system moving together with the periodic orbit. If the deterministic energy of the periodic orbit coincides with

an eigenenergy E of the local Schrödinger equation around the periodic orbit, then the corresponding stationary eigenfunction fulfills

$$\psi_p(\xi, t + T_p) = \int d\xi' G_p(\xi, \xi', t + T_p) \psi_p(\xi', t) = e^{-iET_p/\hbar} \psi_p(\xi, t), \quad (\text{A39.35})$$

where T_p is the period of the prime orbit p . If the deterministic energy of the periodic orbit is not an eigenenergy of the local Schrödinger equation, the non-stationary eigenfunctions fulfill

$$\psi_p^\ell(\xi, t + T_p) = \int d\xi' G_p(\xi, \xi', t + T_p) \psi_p^\ell(\xi', t) = e^{-iET_p/\hbar} \lambda_p^\ell(E) \psi_p^\ell(t), \quad (\text{A39.36})$$

where $\ell = (\ell_1, \ell_2, \dots)$ is a multi-index of the possible quantum numbers of the local Schrödinger equation. If the eigenvalues $\lambda_p^\ell(E)$ are known the local functional determinant can be written as

$$\Delta_p(E) = \prod_\ell (1 - \lambda_p^\ell(E)), \quad (\text{A39.37})$$

since $\Delta_p(E)$ is zero at the eigenenergies of the local Schrödinger problem. We can insert the ansatz (A39.28) and reformulate (A39.36) as

$$e^{\frac{i}{\hbar}S(t+T_p)} \varphi_p^\ell(t + T_p) = e^{-iET_p/\hbar} \lambda_p^\ell(E) e^{\frac{i}{\hbar}S(t)} \varphi_p^\ell(t). \quad (\text{A39.38})$$

The phase change is given by the action integral for one period $S(t + T_p) - S(t) = \int_0^{T_p} L(t) dt$. Using this and the identity for the action $S_p(E)$ of the periodic orbit

$$S_p(E) = \oint p dq = \int_0^{T_p} L(t) dt + ET_p, \quad (\text{A39.39})$$

we get

$$e^{\frac{i}{\hbar}S_p(E)} \varphi_p^\ell(t + T_p) = \lambda_p^\ell(E) \varphi_p^\ell(t). \quad (\text{A39.40})$$

Introducing the eigenequation for the amplitude

$$\varphi_p^\ell(t + T_p) = R_{\ell,p}(E) \varphi_p^\ell(t), \quad (\text{A39.41})$$

the local spectral determinant can be expressed as a product for the quantum numbers of the local problem,

$$\Delta_p(E) = \prod_\ell \left(1 - R_{\ell,p}(E) e^{\frac{i}{\hbar}S_p(E)}\right).$$

Since $\hbar$ is a small parameter we can develop a perturbation series for the amplitudes

$$\varphi_p^\ell(t) = \sum_{m=0}^{\infty} \left(\frac{i\hbar}{2}\right)^m \varphi_p^{\ell(m)}(t)$$

which can be inserted into the equation (A39.29) and we get an iterative scheme starting with the semiclassical solution $\varphi^{\ell(0)}$:

$$\begin{aligned} \partial_t \varphi^{\ell(0)} + \nabla \varphi^{\ell(0)} \nabla S + \frac{1}{2} \varphi^{\ell(0)} \Delta S &= 0, \\ \partial_t \varphi^{\ell(m+1)} + \nabla \varphi^{\ell(m+1)} \nabla S + \frac{1}{2} \varphi^{\ell(m+1)} \Delta S &= \Delta \varphi^{\ell(m)}. \end{aligned} \quad (\text{A39.42})$$

The eigenvalue can also be expanded in powers of $i\hbar/2$:

$$R_{\ell,p}(E) = \exp \left\{ \sum_{m=0}^{\infty} \left(\frac{i\hbar}{2} \right)^m C_{\ell,p}^{(m)} \right\} = e^{C_{\ell,p}^{(0)}} \left\{ 1 + \frac{i\hbar}{2} C_{\ell,p}^{(1)} + \left(\frac{i\hbar}{2} \right)^2 \left(\frac{1}{2} (C_{\ell,p}^{(1)})^2 + C_{\ell,p}^{(2)} \right) + \dots \right\}. \quad (\text{A39.43})$$

The eigenvalue equation (A39.41) in $\hbar$ expanded form reads as

$$\begin{aligned} \varphi_p^{\ell(0)}(t + T_p) &= e^{C_{\ell,p}^{(0)}} \varphi_p^{\ell(0)}(t) \\ \varphi_p^{\ell(1)}(t + T_p) &= e^{C_{\ell,p}^{(0)}} \left[\varphi_p^{\ell(1)}(t) + C_{\ell,p}^{(1)} \varphi_p^{\ell(0)}(t) \right] \\ \varphi_p^{\ell(2)}(t + T_p) &= e^{C_{\ell,p}^{(0)}} \left[\varphi_p^{\ell(2)}(t) + C_{\ell,p}^{(1)} \varphi_p^{\ell(1)}(t) + \left(C_{\ell,p}^{(2)} + \frac{1}{2} (C_{\ell,p}^{(1)})^2 \right) \varphi_p^{\ell(0)}(t) \right], \end{aligned}$$

and so on. These equations are the conditions selecting the eigenvectors and eigenvalues and they hold for all t .

It is convenient to expand the functions $\varphi_p^{\ell(m)}(x, t)$ in Taylor series around the periodic orbit and to solve the equations (A39.43) in this basis [9], since only a couple of coefficients should be computed to derive the first corrections. One can derive in general the zero order term $C_{\ell}^{(0)} = i\pi\nu_p + \sum_{i=1}^{D-1} \left(\ell_i + \frac{1}{2} \right) u_{p,i}$, where $u_{p,i} = \log \Lambda_{p,i}$ are the logarithms of the eigenvalues of the monodromy matrix M_p and ν_p is the topological index of the periodic orbit. The first correction is given by the integral

$$C_{\ell,p}^{(1)} = \int_0^{T_p} dt \frac{\Delta \varphi_p^{\ell(0)}(t)}{\varphi_p^{\ell(0)}(t)}.$$

When the theory is applied for billiard systems, the wave function should fulfill the Dirichlet boundary condition on hard walls, e.g. it should vanish on the wall. The wave function determined from (A39.29) behaves discontinuously when the trajectory $x_p(t)$ hits the wall. For the simplicity we consider a 2-degrees of freedom billiard system here. The wave function on the wall before the bounce (t_{-0}) is given by

$$\psi_{in}(x, y(x), t) = \varphi(x, y(x), t_{-0}) e^{iS(x, y(x), t_{-0})/\hbar}, \quad (\text{A39.44})$$

where $y(x) = Y_2 x^2/2! + Y_3 x^3/3! + Y_4 x^4/4! + \dots$ is the parametrization of the wall around the point of reflection (see Fig 1.). The wave function on the wall after the bounce (t_{+0}) is

$$\psi_{out}(x, y(x), t) = \varphi(x, y(x), t_{+0}) e^{iS(x, y(x), t_{+0})/\hbar}. \quad (\text{A39.45})$$

The sum of these wave functions should vanish on the hard wall. This implies that the incoming and the outgoing amplitudes and the phases are related as

$$S(x, y(x), t_{-0}) = S(x, y(x), t_{+0}), \quad (\text{A39.46})$$

and

$$\varphi(x, y(x), t_{-0}) = -\varphi(x, y(x), t_{+0}). \quad (\text{A39.47})$$

The minus sign can be interpreted as the topological phase coming from the hard wall.

Now we can reexpress the spectral determinant with the local eigenvalues:

$$\Delta(E) = e^{W(E)} \prod_p \prod_{\ell} \left(1 - R_{\ell,p}(E) e^{\frac{i}{\hbar} S_p(E)}\right). \quad (\text{A39.48})$$

This expression is the quantum generalization of the semiclassical Selberg-product formula [19]. A similar decomposition has been found for quantum baker maps in refs. [18, 20]. The functions

$$\zeta_{\ell}^{-1}(E) = \prod_p \left(1 - R_{\ell,p}(E) e^{\frac{i}{\hbar} S_p(E)}\right) \quad (\text{A39.49})$$

are the generalizations of the Ruelle type [17] zeta functions. The trace formula can be recovered from (A39.33):

$$\text{Tr } G(E) = g_0(E) + \frac{1}{i\hbar} \sum_{p,\ell} \left(T_p(E) - i\hbar \frac{d \log R_{\ell,p}(E)}{dE} \right) \frac{R_{\ell,p}(E) e^{\frac{i}{\hbar} S_p(E)}}{1 - R_{\ell,p}(E) e^{\frac{i}{\hbar} S_p(E)}}. \quad (\text{A39.50})$$

We can rewrite the denominator as a sum of a geometric series and we get

$$\text{Tr } G(E) = g_0(E) + \frac{1}{i\hbar} \sum_{p,r,\ell} \left(T_p(E) - i\hbar \frac{d \log R_{\ell,p}(E)}{dE} \right) (R_{\ell,p}(E))^r e^{\frac{i}{\hbar} r S_p(E)}. \quad (\text{A39.51})$$

The new index r can be interpreted as the repetition number of the prime orbit p . This expression is the generalization of the semiclassical trace formula for the exact quantum mechanics. We would like to stress here, that the perturbation calculus introduced above is just one way to compute the eigenvalues of the local Schrödinger problems. Non-perturbative methods can be used to calculate the local eigenvalues for stable, unstable and marginal orbits. Therefore, our trace formula is not limited to integrable or hyperbolic systems, it can describe the most general case of systems with mixed phase space.

The semiclassical trace formula can be recovered by dropping the sub-leading term $-i\hbar d \log R_{\ell,p}(E)/dE$ and using the semiclassical eigenvalue $R_{\ell,p}^{(0)}(E) = e^{C_p^{(\ell(0))}} = e^{-i\nu_p \pi} e^{-\sum_i (\ell_i + 1/2) u_{p,i}}$. Summation for the indexes ℓ_i yields the celebrated semiclassical amplitude

$$\sum_{\ell} (R_{\ell,p}^{(0)}(E))^r = \frac{e^{-ir\nu_p \pi}}{|\det(\mathbf{1} - M_p^r)|^{1/2}}. \quad (\text{A39.52})$$



example A39.1
p. 1117

Commentary

Remark A39.1. Birkhoff normal form. Normal forms of equilibria stability manifolds for 1D Hamiltonian (are-preserving) systems were introduced by Birkhoff [6]. Their convergence is discussed by Moser [16]. In 1927 Birkhoff [7] extended the concept of normal form to give successive integrable approximations to a non-integrable problem.

References

- [1] D. Alonso and P. Gaspard, “ $\hbar$ expansion for the periodic-orbit quantization of chaotic systems”, *Chaos* **3**, 601–612 (1993).
- [2] V. M. Babič and V. S. Buldyrev, *Short-Wavelength Diffraction Theory: Asymptotic Methods* (Springer, New York, 1991).
- [3] R. Balian and C. Bloch, “Distribution of eigenfrequencies for the wave equation in a finite domain: I. Three-dimensional problem with smooth boundary surface”, *Ann. Phys.* **60**, 401–447 (1970).
- [4] R. Balian and C. Bloch, “Distribution of eigenfrequencies for the wave equation in a finite domain. II. Electromagnetic field. Riemannian spaces”, *Ann. Phys.* **64**, 271–307 (1971).
- [5] M. V. Berry and C. J. Howls, “High orders of the Weyl expansion for quantum billiards: resurgence of periodic orbits, and the Stokes phenomenon”, *Proc. R. Soc. Lond. A* **447**, 527–555 (1994).
- [6] G. D. Birkhoff, “Surface transformations and their dynamical applications”, *Acta Math.* **43**, 1–119 (1922).
- [7] G. D. Birkhoff, “On the periodic motions of dynamical systems”, *Acta Math.* **50**, Reprinted in ref. [14], 359–379 (1927).
- [8] P. Cvitanović and B. Eckhardt, “Periodic orbit quantization of chaotic systems”, *Phys. Rev. Lett.* **63**, 823–826 (1989).
- [9] P. Cvitanović, P. E. Rosenqvist, G. Vattay, and H. H. Rugh, “A Fredholm determinant for semiclassical quantization”, *Chaos* **3**, 619–636 (1993).
- [10] P. Gaspard and D. Alonso, “ $\hbar$ expansion for the periodic-orbit quantization of hyperbolic systems”, *Phys. Rev. A* **47**, R3468–R3471 (1993).
- [11] P. Gaspard and S. A. Rice, “Exact quantization of the scattering from a classically chaotic repeller”, *J. Chem. Phys.* **90**, 2255–2262 (1989).
- [12] P. Gaspard and S. A. Rice, “Scattering from a classically chaotic repeller”, *J. Chem. Phys.* **90**, 2225–2241 (1989).
- [13] P. Gaspard and S. A. Rice, “Semiclassical quantization of the scattering from a classically chaotic repeller”, *J. Chem. Phys.* **90**, 2242–2254 (1989).
- [14] R. S. MacKay and J. D. Meiss, *Hamiltonian Dynamical Systems: A Reprint Selection* (Taylor & Francis, Oxford, 1987).

- [15] V. P. Maslov and M. V. Fedoriuk, *Semi-Classical Approximation in Quantum Mechanics* (Reidel, Boston, 1981).
- [16] J. Moser, “The analytic invariants of an area-preserving mapping near a hyperbolic fixed point”, *Commun. Pure Appl. Math.* **9**, 673–692 (1956).
- [17] D. Ruelle, “Resonances for Axiom A flows”, *J. Diff. Geom.* **25**, 99–116 (1987).
- [18] M. Saraceno and A. Voros, “Towards a semiclassical theory of the quantum baker’s map”, *Physica D* **79**, 206–268 (1994).
- [19] A. Voros, “Unstable periodic orbits and semiclassical quantisation”, *J. Phys. A* **21**, 685 (1988).
- [20] A. Voros, “Aspects of semiclassical theory in the presence of classical chaos”, *Progr. Theor. Phys. Suppl.* **116**, 17–44 (1994).
- [21] A. Wirzba, “Validity of the semiclassical periodic orbit approximation in the two- and three-disk problems”, *Chaos* **2**, 77–83 (1992).
- [22] A. Wirzba, “Test of the periodic-orbit approximation in n-disk systems”, *Nucl. Phys. A* **560**, 136–150 (1993).

A39.5 Examples

Example A39.1. The first correction $C_{p,\ell}^{(1)}$ to 2-degrees of freedom billiards. To have an impression about the improvement caused by the quantum corrections we have developed a numerical code which calculates the first correction $C_{p,\ell}^{(1)}$ for general 2-degrees of freedom billiard systems. The first correction depends only on some basic data of the periodic orbit such as the lengths of the free flights between bounces, the angles of incidence and the first three Taylor expansion coefficients Y_2, Y_3, Y_4 of the wall in the point of incidence. To check that our new local method gives the same result as the direct calculation of the Feynman integral, we computed the first $\hbar$ correction $C_{p,0}^{(1)}$ for the periodic orbits of the 3-disk scattering system [11–13] where the quantum corrections have been found to agree up to the fifth decimal digit, while our method generates these numbers with any desired precision. Unfortunately, the $\ell \neq 0$ coefficients cannot be compared to ref. [1], since the ℓ dependence was not realized there due to the lack of general formulas (A39.48) and (A39.49). However, the ℓ dependence can be checked on the 2 disk scattering system [21, 22]. On the standard example [1, 8, 10–13, 21, 22], when R , the distance of the centers, is 6 times the disk radius a , we obtain

$$C_{\ell}^{(1)} = \frac{1}{\sqrt{2E}}(-0.625 \ell^3 - 0.3125 \ell^2 + 1.4375 \ell + 0.625).$$

For $\ell = 0$ and 1 this has been confirmed by A. Wirzba, who was able to compute $C_0^{(1)}$ from his exact quantum calculation. Our method makes it possible to utilize the Cvitanović–Eckhardt [8] symmetry reduction and to repeat the fundamental domain cycle expansion calculation of ref. [8] with the first quantum correction. We computed the correction to the leading 226 prime periodic orbits with 10 or less bounces in the fundamental domain. Table A39.1 shows the numerical values of the exact quantum calculation [21, 22], the semiclassical cycle expansion [9] and our corrected calculation. The error of the corrected calculation vs. the error of the semiclassical calculation decreases with the

Figure A39.2: A typical bounce off a billiard wall. The wall can be characterized by the local expansion $y(x) = Y_2x^2/2! + Y_3x^3/3! + Y_4x^4/4! + \dots$

Table A39.1: Real part of the resonances ($\text{Re } k$) of the 3-disk scattering system at disk separation 6:1. Semiclassical and first corrected cycle expansion versus exact quantum calculation and the error of the semiclassical δ_{SC} divided by the error of the first correction δ_{Corr} . The magnitude of the error in the imaginary part of the resonances remains unchanged.

Quantum	Semiclassical	First correction	$\delta_{SC}/\delta_{Corr}$
0.697995	0.758313	0.585150	0.53
2.239601	2.274278	2.222930	2.08
3.762686	3.787876	3.756594	4.13
5.275666	5.296067	5.272627	6.71
6.776066	6.793636	6.774061	8.76
...	...	...	...
30.24130	30.24555	30.24125	92.3
31.72739	31.73148	31.72734	83.8
32.30110	32.30391	32.30095	20.0
33.21053	33.21446	33.21048	79.4
33.85222	33.85493	33.85211	25.2
34.69157	34.69534	34.69152	77.0

wavenumber. Besides the improved results, a convergence up to six decimal digits is observed, in contrast to just three decimal digits obtained in the full domain calculation [1, 10].

[return to p. 1115](#)

Appendix A40

Infinite dimensional operators

What is a matrix?

—Werner Heisenberg (1925)

What is The Matrix?

—Keanu Reeves (1999)

(A. Wirzba)

THIS APPENDIX, taken from ref. [11], summarizes the definitions and properties of trace-class and Hilbert-Schmidt matrices, the determinants over infinite dimensional matrices and regularization schemes for matrices or operators which are not of trace-class.

Why should a working physicist care about linear algebra? Physicists were blissfully ignorant of it until 1920's, but with Heisenberg's sojourn in Helgoland, everything changed. Quantum Mechanics was formulated as

$$\phi(t) = \hat{U}^t \phi(0), \quad \hat{U}^t = e^{-\frac{i}{\hbar} t \hat{H}}, \quad (\text{A40.1})$$

where $\phi(t)$ is the quantum wave function t , $\hat{U}^t$ is the unitary quantum evolution operator, and $\hat{H}$ is the Hamiltonian operator.

How are we to think of the quantum operator

$$\hat{H} = \hat{T} + \hat{V}, \quad \hat{T} = \hat{p}^2/2m, \quad \hat{V} = V(\hat{q}), \quad (\text{A40.2})$$

corresponding to a deterministic Hamiltonian $H = T + V$, where T is kinetic energy, and V is the potential? What does equation (A40.1) *mean*? In sect. 4.4 we have deconstructed it, by making $\hat{U}^t$ computationally explicit as a the time-ordered product (4.19).

Whenever you are confused about an “operator”, think “matrix”. Expressed in terms of basis functions, the quantum evolution operator is an infinite-dimensional

matrix; if we happen to know the eigenbasis of the Hamiltonian, the problem is solved already. The matrices that have to be evaluated are very high-dimensional, in principle infinite dimensional, and the numerical challenges can quickly get out of hand. What made it possible to solve these equations analytically in 1920's for a few iconic problems, such as the hydrogen atom, are the symmetries, or in other words group theory. In real life we have to guess that some complete basis set is good starting point for solving the problem, and go from there. In practice we truncate such operator representations to finite-dimensional matrices, so it pays to recapitulate a few relevant facts about matrix algebra and some of the properties of functions of finite-dimensional matrices.

A40.1 Matrix-valued functions

(P. Cvitanović)

As a preliminary we summarize some of the properties of functions of finite-dimensional matrices.

The derivative of a matrix is a matrix with elements

$$A'(x) = \frac{dA(x)}{dx}, \quad A'_{ij}(x) = \frac{d}{dx} A_{ij}(x). \quad (\text{A40.3})$$

Derivatives of products of matrices are evaluated by the chain rule

$$\frac{d}{dx}(AB) = \frac{dA}{dx}B + A\frac{dB}{dx}. \quad (\text{A40.4})$$

A matrix and its derivative matrix in general do not commute

$$\frac{d}{dx}A^2 = \frac{dA}{dx}A + A\frac{dA}{dx}. \quad (\text{A40.5})$$

The derivative of the inverse of a matrix, if the inverse exists, follows from $\frac{d}{dx}(AA^{-1}) = 0$:

$$\frac{d}{dx}A^{-1} = -\frac{1}{A}\frac{dA}{dx}\frac{1}{A}. \quad (\text{A40.6})$$

A function of a single variable that can be expressed in terms of additions and multiplications generalizes to a matrix-valued function by replacing the variable by the matrix.

In particular, the exponential of a constant matrix can be defined either by its series expansion, or as a limit of an infinite product:

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k, \quad A^0 = \mathbf{1} \quad (\text{A40.7})$$

$$= \lim_{N \rightarrow \infty} \left(\mathbf{1} + \frac{1}{N} A \right)^N \quad (\text{A40.8})$$

The first equation follows from the second one by the binomial theorem, so these indeed are equivalent definitions. That the terms of order $O(N^{-2})$ or smaller do not matter follows from the bound

$$\left(1 + \frac{x - \epsilon}{N}\right)^N < \left(1 + \frac{x + \delta x_N}{N}\right)^N < \left(1 + \frac{x + \epsilon}{N}\right)^N,$$

where $|\delta x_N| < \epsilon$. If $\lim \delta x_N \rightarrow 0$ as $N \rightarrow \infty$, the extra terms do not contribute.

Consider now the determinant

$$\det(e^A) = \lim_{N \rightarrow \infty} (\det(\mathbf{1} + A/N))^N.$$

To the leading order in $1/N$

$$\det(\mathbf{1} + A/N) = 1 + \frac{1}{N} \text{tr } A + O(N^{-2}).$$

hence

$$\det e^A = \lim_{N \rightarrow \infty} \left(1 + \frac{1}{N} \text{tr } A + O(N^{-2})\right)^N = e^{\text{tr } A} \quad (\text{A40.9})$$

Due to non-commutativity of matrices, generalization of a function of several variables to a function of several matrices is not as straightforward. Expression involving several matrices depend on their commutation relations. For example, the Baker-Campbell-Hausdorff commutator expansion

$$e^{tA} B e^{-tA} = B + t[A, B] + \frac{t^2}{2}[A, [A, B]] + \frac{t^3}{3!}[A, [A, [A, B]]] + \dots \quad (\text{A40.10})$$

sometimes used to establish the equivalence of the Heisenberg and Schrödinger pictures of quantum mechanics follows by recursive evaluation of t derivatives

$$\frac{d}{dt} (e^{tA} B e^{-tA}) = e^{tA} [A, B] e^{-tA}.$$

Expanding $\exp(A + B)$, $\exp A$, $\exp B$ to first few orders using (A40.7) yields

$$e^{(A+B)/N} = e^{A/N} e^{B/N} - \frac{1}{2N^2} [A, B] + O(N^{-3}), \quad (\text{A40.11})$$

and the Trotter product formula: if B , C and $A = B + C$ are matrices, then

$$e^A = \lim_{N \rightarrow \infty} (e^{B/N} e^{C/N})^N \quad (\text{A40.12})$$

A40.2 Operator norms

(R. Mainieri and P. Cvitanović)



The limit used in the above definition involves matrices - operators in vector spaces - rather than numbers, and its convergence can be checked using

tools familiar from calculus. We briefly review those tools here, as throughout the text we will have to consider many different operators and how they converge.

The $n \rightarrow \infty$ convergence of partial products

$$\mathbf{E}_n = \prod_{0 \leq m < n} \left(\mathbf{1} + \frac{t}{m} \mathbf{A} \right)$$

can be verified using the Cauchy criterion, which states that the sequence $\{\mathbf{E}_n\}$ converges if the differences $\|\mathbf{E}_k - \mathbf{E}_j\| \rightarrow 0$ as $k, j \rightarrow \infty$. To make sense of this we need to define a sensible norm $\|\cdot\|$. Norm of a matrix is based on the Euclidean norm for a vector: the idea is to assign to a matrix $\mathbf{M}$ a norm that is the largest possible change it can cause to the length of a unit vector $\hat{n}$:

$$\|\mathbf{M}\| = \sup_{\hat{n}} \|\mathbf{M}\hat{n}\|, \quad \|\hat{n}\| = 1. \quad (\text{A40.13})$$

We say that $\|\cdot\|$ is the operator norm induced by the vector norm $\|\cdot\|$. Constructing a norm for a finite-dimensional matrix is easy, but had $\mathbf{M}$ been an operator in an infinite-dimensional space, we would also have to specify the space $\hat{n}$ belongs to. In the finite-dimensional case, the sum of the absolute values of the components of a vector is also a norm; the induced operator norm for a matrix $\mathbf{M}$ with components M_{ij} in that case can be defined by

$$\|\mathbf{M}\| = \max_i \sum_j |M_{ij}|. \quad (\text{A40.14})$$

The operator norm (A40.14) and the vector norm (A40.13) are only rarely distinguished by different notation, a bit of notational laziness that we shall uphold.

Now that we have learned how to make sense out of norms of operators, we can check that

$$\|e^{t\mathbf{A}}\| \leq e^{t\|\mathbf{A}\|}. \quad (\text{A40.15})$$

exercise A40.1

As $\|\mathbf{A}\|$ is a number, the norm of $e^{t\mathbf{A}}$ is finite and therefore well defined. In particular, the exponential of a matrix is well defined for all values of t , and the linear differential equation (4.11) has a solution for all times.

exercise 2.9

A40.3 Trace class and Hilbert-Schmidt class

This section is mainly an extract from ref. [8]. Refs. [1, 3, 7, 9] should be consulted for more details and proofs. The trace class and Hilbert-Schmidt property will be defined here for linear, in general non-hermitian operators $\mathbf{A} \in \mathcal{L}(\mathcal{H})$: $\mathcal{H} \rightarrow \mathcal{H}$ (where $\mathcal{H}$ is a separable Hilbert space). The transcription to matrix elements (used in the prior chapters) is simply $a_{ij} = \langle \phi_i, \mathbf{A}\phi_j \rangle$ where $\{\phi_n\}$ is an orthonormal basis of $\mathcal{H}$ and $\langle, \rangle$ is the inner product in $\mathcal{H}$ (see sect. A40.5 where the theory of *von Koch matrices* of ref. [5] is discussed). So, the trace is the generalization of the usual notion of the sum of the diagonal elements of a matrix; but because infinite sums are involved, not all operators will have a trace:

Definition:

- (a) An operator $\mathbf{A}$ is called **trace class**, $\mathbf{A} \in \mathcal{J}_1$, if and only if, for every orthonormal basis, $\{\phi_n\}$:

$$\sum_n |\langle \phi_n, \mathbf{A} \phi_n \rangle| < \infty. \quad (\text{A40.16})$$

The family of all trace class operators is denoted by $\mathcal{J}_1$.

- (b) An operator $\mathbf{A}$ is called **Hilbert-Schmidt**, $\mathbf{A} \in \mathcal{J}_2$, if and only if, for every orthonormal basis, $\{\phi_n\}$:

$$\sum_n \|\mathbf{A} \phi_n\|^2 < \infty.$$

The family of all Hilbert-Schmidt operators is denoted by $\mathcal{J}_2$.

Bounded operators are dual to trace class operators. They satisfy the following condition: $|\langle \psi, B\phi \rangle| \leq C \|\psi\| \|\phi\|$ with $C < \infty$ and $\psi, \phi \in \mathcal{H}$. If they have eigenvalues, these are bounded too. The family of bounded operators is denoted by $\mathcal{B}(\mathcal{H})$ with the norm $\|B\| = \sup_{\phi \neq 0} \frac{\|B\phi\|}{\|\phi\|}$ for $\phi \in \mathcal{H}$. Examples for bounded operators are unitary operators and especially the unit matrix. In fact, every bounded operator can be written as linear combination of four unitary operators.

A bounded operator $\mathbf{C}$ is *compact*, if it is the norm limit of finite rank operators.

An operator $\mathbf{A}$ is called *positive*, $\mathbf{A} \geq 0$, if $\langle \mathbf{A} \phi, \phi \rangle \geq 0 \ \forall \phi \in \mathcal{H}$. Note that $\mathbf{A}^\dagger \mathbf{A} \geq 0$. We define $|\mathbf{A}| = \sqrt{\mathbf{A}^\dagger \mathbf{A}}$.

The most important properties of the trace and Hilbert-Schmidt classes are summarized in (see refs. [7, 8]):

- (a) $\mathcal{J}_1$ and $\mathcal{J}_2$ are **ideals*, i.e., they are vector spaces closed under scalar multiplication, sums, adjoints, and multiplication with bounded operators.
- (b) $\mathbf{A} \in \mathcal{J}_1$ if and only if $\mathbf{A} = \mathbf{B}\mathbf{C}$ with $\mathbf{B}, \mathbf{C} \in \mathcal{J}_2$.
- (c) $\mathcal{J}_1 \subset \mathcal{J}_2 \subset \text{Compact operators}$.
- (d) For any operator $\mathbf{A}$, we have $\mathbf{A} \in \mathcal{J}_2$ if $\sum_n \|\mathbf{A} \phi_n\|^2 < \infty$ for a single basis. For any operator $\mathbf{A} \geq 0$ we have $\mathbf{A} \in \mathcal{J}_1$ if $\sum_n |\langle \phi_n, \mathbf{A} \phi_n \rangle| < \infty$ for a single basis.
- (e) If $\mathbf{A} \in \mathcal{J}_1$, $\text{Tr}(\mathbf{A}) = \sum \langle \phi_n, \mathbf{A} \phi_n \rangle$ is independent of the basis used.
- (f) Tr is linear and obeys $\text{Tr}(\mathbf{A}^\dagger) = \overline{\text{Tr}(\mathbf{A})}$; $\text{Tr}(\mathbf{A}\mathbf{B}) = \text{Tr}(\mathbf{B}\mathbf{A})$ if either $\mathbf{A} \in \mathcal{J}_1$ and $\mathbf{B}$ bounded, $\mathbf{A}$ bounded and $\mathbf{B} \in \mathcal{J}_1$ or both $\mathbf{A}, \mathbf{B} \in \mathcal{J}_2$.
- (g) $\mathcal{J}_2$ endowed with the inner product $\langle \mathbf{A}, \mathbf{B} \rangle_2 = \text{Tr}(\mathbf{A}^\dagger \mathbf{B})$ is a Hilbert space. If $\|\mathbf{A}\|_2 = [\text{Tr}(\mathbf{A}^\dagger \mathbf{A})]^{1/2}$, then $\|\mathbf{A}\|_2 \geq \|\mathbf{A}\|$ and $\mathcal{J}_2$ is the $\|\cdot\|_2$ -closure of the *finite* rank operators.

- (h) $\mathcal{J}_1$ endowed with the norm $\|\mathbf{A}\|_1 = \text{Tr}(\sqrt{\mathbf{A}^\dagger \mathbf{A}})$ is a Banach space. $\|\mathbf{A}\|_1 \geq \|\mathbf{A}\|_2 \geq \|\mathbf{A}\|$ and $\mathcal{J}_1$ is the $\|\cdot\|_1$ -norm closure of the *finite* rank operators. The dual space of $\mathcal{J}_1$ is $\mathcal{B}(\mathcal{H})$, the family of bounded operators with the duality $\langle \mathbf{B}, \mathbf{A} \rangle = \text{Tr}(\mathbf{B}\mathbf{A})$.
- (i) If $\mathbf{A}, \mathbf{B} \in \mathcal{J}_2$, then $\|\mathbf{AB}\|_1 \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_2$. If $\mathbf{A} \in \mathcal{J}_2$ and $\mathbf{B} \in \mathcal{B}(\mathcal{H})$, then $\|\mathbf{AB}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|$. If $\mathbf{A} \in \mathcal{J}_1$ and $\mathbf{B} \in \mathcal{B}(\mathcal{H})$, then $\|\mathbf{AB}\|_1 \leq \|\mathbf{A}\|_1 \|\mathbf{B}\|$.

Note the most important property for proving that an operator is trace class is the decomposition (b) into two Hilbert-Schmidt ones, as the Hilbert-Schmidt property can easily be verified in one single orthonormal basis (see (d)). Property (e) ensures then that the trace is the same in any basis. Properties (a) and (f) show that trace class operators behave in complete analogy to finite rank operators. The proof whether a matrix is trace-class (or Hilbert-Schmidt) or not simplifies enormously for diagonal matrices, as then the second part of property (d) is directly applicable: just the moduli of the eigenvalues (or – in case of Hilbert-Schmidt – the squares of the eigenvalues) have to be summed up in order to answer that question. A good strategy in checking the trace-class character of a general matrix $\mathbf{A}$ is therefore the decomposition of that matrix into two matrices $\mathbf{B}$ and $\mathbf{C}$ where one, say $\mathbf{C}$, should be chosen to be diagonal and either just barely of Hilbert-Schmidt character leaving enough freedom for its partner $\mathbf{B}$ or of trace-class character such that one only has to show the boundedness for $\mathbf{B}$.

A40.4 Determinants of trace class operators

This section is mainly based on refs. [6, 9] which should be consulted for more details and proofs. See also refs. [1, 3].

Pre-definitions (Alternating algebra and Fock spaces):

Given a Hilbert space $\mathcal{H}$, $\otimes^n \mathcal{H}$ is defined as the vector space of multi-linear functionals on $\mathcal{H}$ with $\phi_1 \otimes \cdots \otimes \phi_n \in \otimes^n \mathcal{H}$ in case $\phi_1, \dots, \phi_n \in \mathcal{H}$. $\wedge^n(\mathcal{H})$ is defined as the subspace of $\otimes^n \mathcal{H}$ spanned by the wedge-product

$$\phi_1 \wedge \cdots \wedge \phi_n = \frac{1}{\sqrt{n!}} \sum_{\pi \in \mathcal{P}_n} \epsilon(\pi) [\phi_{\pi(1)} \otimes \cdots \otimes \phi_{\pi(n)}]$$

where $\mathcal{P}_n$ is the group of all permutations of n letters and $\epsilon(\pi) = \pm 1$ depending on whether π is an even or odd permutation, respectively. The inner product in $\wedge^n(\mathcal{H})$ is given by

$$(\phi_1 \wedge \cdots \wedge \phi_n, \eta_1 \wedge \cdots \wedge \eta_n) = \det \{(\phi_i, \eta_j)\}$$

where $\det\{a_{ij}\} = \sum_{\pi \in \mathcal{P}_n} \epsilon(\pi) a_{1\pi(1)} \cdots a_{n\pi(n)}$. $\wedge^n(\mathbf{A})$ is defined as functor (a functor satisfies $\wedge^n(\mathbf{AB}) = \wedge^n(\mathbf{A}) \wedge^n(\mathbf{B})$) on $\wedge^n(\mathcal{H})$ with

$$\wedge^n(\mathbf{A})(\phi_1 \wedge \cdots \wedge \phi_n) = \mathbf{A}\phi_1 \wedge \cdots \wedge \mathbf{A}\phi_n.$$

When $n = 0$, $\wedge^n(\mathcal{H})$ is defined to be C and $\wedge^n(\mathbf{A})$ as $1: C \rightarrow C$.

Properties: If $\mathbf{A}$ trace class, i.e., $\mathbf{A} \in \mathcal{J}_1$, then for any k , $\bigwedge^k(\mathbf{A})$ is trace class, and for any orthonormal basis $\{\phi_n\}$ the cumulant

$$\mathrm{Tr}\left(\bigwedge^k(\mathbf{A})\right) = \sum_{i_1 < \dots < i_k} ((\phi_{i_1} \wedge \dots \wedge \phi_{i_k}), (\mathbf{A}\phi_{i_1} \wedge \dots \wedge \mathbf{A}\phi_{i_k})) < \infty$$

is independent of the basis (with the understanding that $\mathrm{Tr}\bigwedge^0(\mathbf{A}) \equiv 1$).

Definition: Let $\mathbf{A} \in \mathcal{J}_1$, then $\det(1 + \mathbf{A})$ is defined as

$$\det(\mathbf{1} + \mathbf{A}) = \sum_{k=0}^{\infty} \mathrm{Tr}\left(\bigwedge^k(\mathbf{A})\right) \quad (\text{A40.17})$$

Properties:

Let $\mathbf{A}$ be a linear operator on a separable Hilbert space $\mathcal{H}$ and $\{\phi_j\}_1^\infty$ an orthonormal basis.

- (a) $\sum_{k=0}^{\infty} \mathrm{Tr}\left(\bigwedge^k(\mathbf{A})\right)$ converges for each $\mathbf{A} \in \mathcal{J}_1$.
- (b) $|\det(\mathbf{1} + \mathbf{A})| \leq \prod_{j=1}^{\infty} (1 + \mu_j(\mathbf{A}))$ where $\mu_j(\mathbf{A})$ are the *singular* values of $\mathbf{A}$, i.e., the eigenvalues of $|\mathbf{A}| = \sqrt{\mathbf{A}^\dagger \mathbf{A}}$.
- (c) $|\det(\mathbf{1} + \mathbf{A})| \leq \exp(\|\mathbf{A}\|_1)$.
- (d) For any $\mathbf{A}_1, \dots, \mathbf{A}_n \in \mathcal{J}_1$, $\langle z_1, \dots, z_n \rangle \mapsto \det(\mathbf{1} + \sum_{i=1}^n z_i \mathbf{A}_i)$ is an entire analytic function.
- (e) If $\mathbf{A}, \mathbf{B} \in \mathcal{J}_1$, then

$$\begin{aligned} \det(\mathbf{1} + \mathbf{A})\det(\mathbf{1} + \mathbf{B}) &= \det(\mathbf{1} + \mathbf{A} + \mathbf{B} + \mathbf{AB}) \\ &= \det((\mathbf{1} + \mathbf{A})(\mathbf{1} + \mathbf{B})) \\ &= \det((\mathbf{1} + \mathbf{B})(\mathbf{1} + \mathbf{A})) . \end{aligned} \quad (\text{A40.18})$$

If $\mathbf{A} \in \mathcal{J}_1$ and $\mathbf{U}$ unitary, then

$$\det(\mathbf{U}^{-1}(\mathbf{1} + \mathbf{A})\mathbf{U}) = \det(\mathbf{1} + \mathbf{U}^{-1}\mathbf{A}\mathbf{U}) = \det(\mathbf{1} + \mathbf{A}) .$$

- (f) If $\mathbf{A} \in \mathcal{J}_1$, then $(\mathbf{1} + \mathbf{A})$ is invertible if and only if $\det(\mathbf{1} + \mathbf{A}) \neq 0$.
- (g) If $\lambda \neq 0$ is an n -times degenerate eigenvalue of $\mathbf{A} \in \mathcal{J}_1$, then $\det(\mathbf{1} + z\mathbf{A})$ has a zero of order n at $z = -1/\lambda$.
- (h) For any ϵ , there is a $C_\epsilon(\mathbf{A})$, depending on $\mathbf{A} \in \mathcal{J}_1$, so that $|\det(\mathbf{1} + z\mathbf{A})| \leq C_\epsilon(\mathbf{A}) \exp(\epsilon|z|)$.

(i) For any $\mathbf{A} \in \mathcal{J}_1$,

$$\det(\mathbf{1} + \mathbf{A}) = \prod_{j=1}^{N(\mathbf{A})} (1 + \lambda_j(\mathbf{A})) \quad (\text{A40.19})$$

where here and in the following $\{\lambda_j(\mathbf{A})\}_{j=1}^{N(\mathbf{A})}$ are the eigenvalues of $\mathbf{A}$ counted with algebraic multiplicity .

(j) *Lidskii's theorem*: For any $\mathbf{A} \in \mathcal{J}_1$,

$$\text{Tr}(\mathbf{A}) = \sum_{j=1}^{N(\mathbf{A})} \lambda_j(\mathbf{A}) < \infty .$$

(k) If $\mathbf{A} \in \mathcal{J}_1$, then

$$\begin{aligned} \text{Tr} \left(\bigwedge^k (\mathbf{A}) \right) &= \sum_{j=1}^{N(\bigwedge^k (\mathbf{A}))} \lambda_j \left(\bigwedge^k (\mathbf{A}) \right) \\ &= \sum_{1 \leq j_1 < \dots < j_k \leq N(\mathbf{A})} \lambda_{j_1}(\mathbf{A}) \cdots \lambda_{j_k}(\mathbf{A}) < \infty. \end{aligned}$$

(l) If $\mathbf{A} \in \mathcal{J}_1$, then

$$\det(1 + z\mathbf{A}) = \sum_{k=0}^{\infty} z^k \sum_{1 \leq j_1 < \dots < j_k \leq N(\mathbf{A})} \lambda_{j_1}(\mathbf{A}) \cdots \lambda_{j_k}(\mathbf{A}) < \infty. \quad (\text{A40.20})$$

(m) If $\mathbf{A} \in \mathcal{J}_1$, then for $|z|$ small (i.e., $|z| \max |\lambda_j(\mathbf{A})| < 1$) the series $\sum_{k=1}^{\infty} z^k \text{Tr}((- \mathbf{A})^k) / k$ converges and

$$\begin{aligned} \det(1 + z\mathbf{A}) &= \exp \left(- \sum_{k=1}^{\infty} \frac{z^k}{k} \text{Tr}((- \mathbf{A})^k) \right) \\ &= \exp(\text{Tr} \ln(\mathbf{1} + z\mathbf{A})) . \end{aligned} \quad (\text{A40.21})$$

(n) *The Plemelj-Smithies formula*: Define $\alpha_m(\mathbf{A})$ for $\mathbf{A} \in \mathcal{J}_1$ by

$$\det(\mathbf{1} + z\mathbf{A}) = \sum_{m=0}^{\infty} z^m \frac{\alpha_m(\mathbf{A})}{m!} . \quad (\text{A40.22})$$

Then $\alpha_m(\mathbf{A})$ is given by the $m \times m$ determinant:

$$\alpha_m(\mathbf{A}) = \begin{vmatrix} \text{Tr}(\mathbf{A}) & m-1 & 0 & \cdots & 0 \\ \text{Tr}(\mathbf{A}^2) & \text{Tr}(\mathbf{A}) & m-2 & \cdots & 0 \\ \text{Tr}(\mathbf{A}^3) & \text{Tr}(\mathbf{A}^2) & \text{Tr}(\mathbf{A}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{Tr}(\mathbf{A}^m) & \text{Tr}(\mathbf{A}^{m-1}) & \text{Tr}(\mathbf{A}^{m-2}) & \cdots & \text{Tr}(\mathbf{A}) \end{vmatrix} \quad (\text{A40.23})$$

with the understanding that $\alpha_0(\mathbf{A}) \equiv 1$ and $\alpha_1(\mathbf{A}) \equiv \text{Tr}(\mathbf{A})$. Thus the cumulants $c_m(\mathbf{A}) \equiv \alpha_m(\mathbf{A})/m!$ satisfy the following recursion relation

$$\begin{aligned} c_m(\mathbf{A}) &= \frac{1}{m} \sum_{k=1}^m (-1)^{k+1} c_{m-k}(\mathbf{A}) \text{Tr}(\mathbf{A}^k) \quad \text{for } m \geq 1 \\ c_0(\mathbf{A}) &\equiv 1. \end{aligned} \quad (\text{A40.24})$$

Note that in the context of quantum mechanics formula (A40.22) is the quantum analog to the curvature expansion of the semiclassical zeta function with $\text{Tr}(\mathbf{A}^m)$ corresponding to the sum of all periodic orbits (prime and also repeated ones) of *total* topological length m , i.e., let $c_m(\text{s.c.})$ denote the m^{th} curvature term, then the curvature expansion of the semiclassical zeta function is given by the recursion relation

$$\begin{aligned} c_m(\text{s.c.}) &= \frac{1}{m} \sum_{k=1}^m (-1)^{k+m+1} c_{m-k}(\text{s.c.}) \sum_{\substack{p; r \geq 0 \\ \text{with } [p]r=k}} [p] \frac{t_p(k)^r}{1 - \left(\frac{1}{\Lambda_p}\right)^r} \quad \text{for } m \geq 1 \\ c_0(\text{s.c.}) &\equiv 1. \end{aligned} \quad (\text{A40.25})$$

In fact, in the cumulant expansion (A40.22) as well as in the curvature expansion there are large cancelations involved. Let us order – without loss of generality – the eigenvalues of the operator $\mathbf{A} \in \mathcal{J}_1$ as follows:

$$|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_{i-1}| \geq |\lambda_i| \geq |\lambda_{i+1}| \geq \dots$$

(This is always possible because of $\sum_{i=1}^{N(\mathbf{A})} |\lambda_i| < \infty$.) Then, in the standard (Plemelj-Smithies) cumulant evaluation of the determinant, eq. (A40.22), we have enormous cancelations of big numbers, e.g. at the k^{th} cumulant order ($k > 3$), all the intrinsically large ‘numbers’ $\lambda_1^k, \lambda_1^{k-1} \lambda_2, \dots, \lambda_1^{k-2} \lambda_2 \lambda_3, \dots$ and many more have to cancel out exactly until only $\sum_{1 \leq j_1 < \dots < j_k \leq N(\mathbf{A})} \lambda_{j_1} \dots \lambda_{j_k}$ is finally left over. Algebraically, the fact that there are these large cancelations is of course of no importance. However, if the determinant is calculated numerically, the big cancelations might spoil the result or even the convergence. Now, the curvature expansion of the semiclassical zeta function, as it is known today, *is* the semiclassical approximation to the curvature expansion (unfortunately) in the Plemelj-Smithies form. As the exact quantum mechanical result is approximated semiclassically, the errors introduced in the approximation might lead to big effects as they are done with respect to large quantities which eventually cancel out and not – as it would be of course better – with respect to the small surviving cumulants. Thus it would be very desirable to have a semiclassical analog to the reduced cumulant expansion (A40.20) or even to (A40.19) directly. It might not be possible to find a direct semiclassical analog for the individual eigenvalues λ_j . Thus the direct construction of the semiclassical equivalent to (A40.19) is rather unlikely. However, in order to have a semiclassical ‘cumulant’ summation without large cancelations – see (A40.20) – it would be just sufficient to find the semiclassical analog of each complete cumulant (A40.20) and not of the single eigenvalues. Whether this will eventually be possible is still an open question.

A40.5 Von Koch matrices

Implicitly, many of the above properties are based on the theory of von Koch matrices [1, 2, 5]: An infinite matrix $\mathbf{1} - \mathbf{A} = \|\delta_{jk} - a_{jk}\|_1^\infty$, consisting of complex numbers, is called a matrix with an *absolutely convergent determinant*, if the series $\sum |a_{j_1 k_1} a_{j_2 k_2} \cdots a_{j_n k_n}|$ converges, where the sum extends over all pairs of systems of indices $(j_1, j_2, \dots, j_n)$ and $(k_1, k_2, \dots, k_n)$ which differ from each other only by a permutation, and $j_1 < j_2 < \cdots < j_n$ ($n = 1, 2, \dots$). Then the limit

$$\lim_{n \rightarrow \infty} \det \|\delta_{jk} - a_{jk}\|_1^n = \det(\mathbf{1} - \mathbf{A})$$

exists and is called the determinant of the matrix $\mathbf{1} - \mathbf{A}$. It can be represented in the form

$$\det(\mathbf{1} - \mathbf{A}) = 1 - \sum_{j=1}^{\infty} a_{jj} + \frac{1}{2!} \sum_{j,k=1}^{\infty} \begin{vmatrix} a_{jj} & a_{jk} \\ a_{kj} & a_{kk} \end{vmatrix} - \frac{1}{3!} \sum_{j,k,m=1}^{\infty} \begin{vmatrix} a_{jj} & a_{jk} & a_{jm} \\ a_{kj} & a_{kk} & a_{km} \\ a_{mj} & a_{mk} & a_{mm} \end{vmatrix} + \cdots,$$

where the series on the r.h.s. will remain convergent even if the numbers a_{jk} ($j, k = 1, 2, \dots$) are replaced by their moduli and if all the terms obtained by expanding the determinants are taken with the plus sign. The matrix $\mathbf{1} - \mathbf{A}$ is called *von Koch matrix*, if both conditions

$$\sum_{j=1}^{\infty} |a_{jj}| < \infty, \quad (\text{A40.26})$$

$$\sum_{j,k=1}^{\infty} |a_{jk}|^2 < \infty \quad (\text{A40.27})$$

are fulfilled. Then the following holds (see ref. [1, 2]): (1) Every von Koch matrix has an absolutely convergent determinant. If the elements of a von Koch matrix are functions of some parameter μ ($a_{jk} = a_{jk}(\mu)$, $j, k = 1, 2, \dots$) and both series in the defining condition converge uniformly in the domain of the parameter μ , then as $n \rightarrow \infty$ the determinant $\det \|\delta_{jk} - a_{jk}(\mu)\|_1^n$ tends to the determinant $\det(\mathbf{1} - \mathbf{A}(\mu))$ uniformly with respect to μ , over the domain of μ . (2) If the matrices $\mathbf{1} - \mathbf{A}$ and $\mathbf{1} - \mathbf{B}$ are von Koch matrices, then their product $\mathbf{1} - \mathbf{C} = (\mathbf{1} - \mathbf{A})(\mathbf{1} - \mathbf{B})$ is a von Koch matrix, and

$$\det(\mathbf{1} - \mathbf{C}) = \det(\mathbf{1} - \mathbf{A}) \det(\mathbf{1} - \mathbf{B}).$$

Note that every trace-class matrix $\mathbf{A} \in \mathcal{J}_1$ is also a von Koch matrix (and that any matrix satisfying condition (A40.27) is Hilbert-Schmidt and vice versa). The inverse implication, however, is not true: von Koch matrices are not automatically trace-class. The caveat is that the definition of von Koch matrices is basis-dependent, whereas the trace-class property is basis-independent. As the traces

involve infinite sums, the basis-independence is not at all trivial. An example for an infinite matrix which is von Koch, but not trace-class is the following:

$$\mathbf{A}_{ij} = \begin{cases} 2/j & \text{for } i - j = -1 \text{ and } j \text{ even,} \\ 2/i & \text{for } i - j = +1 \text{ and } i \text{ even,} \\ 0 & \text{else,} \end{cases}$$

i.e.,

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1/2 & 0 & 0 & \cdots \\ 0 & 0 & 1/2 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 1/3 & \ddots \\ 0 & 0 & 0 & 0 & 1/3 & 0 & \ddots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}. \quad (\text{A40.28})$$

Obviously, condition (A40.26) is fulfilled by definition. Second, the condition (A40.27) is satisfied as $\sum_{n=1}^{\infty} 2/n^2 < \infty$. However, the sum over the moduli of the eigenvalues is just twice the harmonic series $\sum_{n=1}^{\infty} 1/n$ which does not converge. The matrix (A40.28) violates the trace-class definition (A40.16), as in its eigenbasis the sum over the moduli of its diagonal elements is infinite. Thus the *absolute* convergence is traded for a *conditional* convergence, since the sum over the eigenvalues themselves can be arranged to still be zero, if the eigenvalues with the same modulus are summed first. Absolute convergence is of course essential, if sums have to be rearranged or exchanged. Thus, the trace-class property is indispensable for any controlled unitary transformation of an infinite determinant, as then there will be necessarily a change of basis and in general also a re-ordering of the corresponding traces. Therefore the claim that a *Hilbert-Schmidt operator with a vanishing trace is automatically trace-class* is false. In general, such an operator has to be regularized in addition (see next chapter).

A40.6 Regularization

Many interesting operators are not of trace class (although they might be in some $\mathcal{J}_p$ with $p > 1$ - an operator A is in $\mathcal{J}_p$ iff $\text{Tr}|A|^p < \infty$ in any orthonormal basis). In order to compute determinants of such operators, an extension of the cumulant expansion is needed which in fact corresponds to a regularization procedure [6, 9]:

E.g. let $\mathbf{A} \in \mathcal{J}_p$ with $p \leq n$. Define

$$R_n(z\mathbf{A}) = (\mathbf{1} + z\mathbf{A}) \exp \left(\sum_{k=1}^{n-1} \frac{(-z)^k}{k} \mathbf{A}^k \right) - \mathbf{1}$$

as the regulated version of the operator $z\mathbf{A}$. Then the regulated operator $R_n(z\mathbf{A})$ is trace class, i.e., $R_n(z\mathbf{A}) \in \mathcal{J}_1$. Define now $\det_n(\mathbf{1} + z\mathbf{A}) = \det(\mathbf{1} + R_n(z\mathbf{A}))$. Then

the regulated determinant

$$\det_n(\mathbf{1} + z\mathbf{A}) = \prod_{j=1}^{N(z\mathbf{A})} \left[\left(1 + z\lambda_j(\mathbf{A}) \right) \exp \left(\sum_{k=1}^{n-1} \frac{(-z\lambda_j(\mathbf{A}))^k}{k} \right) \right] < \infty. \quad (\text{A40.29})$$

exists and is finite. The corresponding Plemelj-Smithies formula now reads [9]:

$$\det_n(\mathbf{1} + z\mathbf{A}) = \sum_{m=0}^{\infty} z^m \frac{\alpha_m^{(n)}(\mathbf{A})}{m!}. \quad (\text{A40.30})$$

with $\alpha_m^{(n)}(\mathbf{A})$ given by the $m \times m$ determinant:

$$\alpha_m^{(n)}(\mathbf{A}) = \begin{vmatrix} \sigma_1^{(n)} & m-1 & 0 & \cdots & 0 \\ \sigma_2^{(n)} & \sigma_1^{(n)} & m-2 & \cdots & 0 \\ \sigma_3^{(n)} & \sigma_2^{(n)} & \sigma_1^{(n)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_m^{(n)} & \sigma_{m-1}^{(n)} & \sigma_{m-2}^{(n)} & \cdots & \sigma_1^{(n)} \end{vmatrix} \quad (\text{A40.31})$$

where

$$\sigma_k^{(n)} = \begin{cases} \text{Tr}(\mathbf{A}^k) & k \geq n \\ 0 & k \leq n-1 \end{cases}$$

As Simon [9] says simply, the beauty of (A40.31) is that we get $\det_n(\mathbf{1} + \mathbf{A})$ from the standard Plemelj-Smithies formula (A40.22) by simply setting $\text{Tr}(\mathbf{A})$, $\text{Tr}(\mathbf{A}^2)$, $\dots$, $\text{Tr}(\mathbf{A}^{n-1})$ to zero.

See also Voros [10] where $\{\lambda_j\}$ are the eigenvalues of an elliptic (pseudo)-differential operator $\mathbf{H}$ of order m on a compact or bounded manifold of dimension d , $0 < \lambda_0 \leq \lambda_1 \leq \dots$ and $\lambda_k \uparrow +\infty$, and the Fredholm determinant

$$\Delta(\lambda) = \prod_{k=0}^{\infty} \left(1 - \frac{\lambda}{\lambda_k} \right)$$

is regularized in the case $\mu \equiv d/m > 1$ as the Weierstrass product

$$\Delta(\lambda) = \prod_{k=0}^{\infty} \left[\left(1 - \frac{\lambda}{\lambda_k} \right) \exp \left(\frac{\lambda}{\lambda_k} + \frac{\lambda^2}{2\lambda_k^2} + \cdots + \frac{\lambda^{[\mu]}}{[\mu]\lambda_k^{[\mu]}} \right) \right] \quad (\text{A40.32})$$

where $[\mu]$ denotes the integer part of μ . This is the unique entire function [10] of order μ having zeros at $\{\lambda_k\}$, and subject to the normalization conditions

$$\ln \Delta(0) = \frac{d}{d\lambda} \ln \Delta(0) = \cdots = \frac{d^{[\mu]}}{d\lambda^{[\mu]}} \ln \Delta(0) = 0.$$

Clearly (A40.32) is the same as (A40.29); one just has to identify $z = -\lambda$, $\mathbf{A} = 1/\mathbf{H}$ and $n-1 = [\mu]$. An example is the regularization of the spectral determinant

$$\Delta(E) = \det[(E - \mathbf{H})] \quad (\text{A40.33})$$

which –as it stands– would only make sense for a finite dimensional basis (or finite dimensional matrices). In ref. [4] the regulated spectral determinant for the example of the hyperbola billiard in two dimensions (thus $d = 2$, $m = 2$ and hence $\mu = 1$) is given as

$$\Delta(E) = \det [(E - \mathbf{H})\Omega(E, \mathbf{H})]$$

where

$$\Omega(E, \mathbf{H}) = -\mathbf{H}^{-1} e^{E\mathbf{H}^{-1}}$$

such that the spectral determinant in the eigenbasis of $\mathbf{H}$ (with eigenvalues $E_n \neq 0$) reads

$$\Delta(E) = \prod_n \left(1 - \frac{E}{E_n}\right) e^{E/E_n} < \infty.$$

Note that $\mathbf{H}^{-1}$ is for this example of Hilbert-Schmidt character.

References

- [1] I. C. Gohberg and M. G. Krein, *Introduction to the Theory of Linear Non-selfadjoint Operators*, Vol. 18, Translations of Mathematical Monographs (Amer. Math. Soc., Providence RI, 1969).
- [2] E. Hille and J. D. Tamarkin, “On the characteristic values of linear integral equations”, *Acta Math.* **57**, 1–76 (1931).
- [3] T. Kato, *Perturbation Theory for Linear Operators*, Chap. X, sects. 1.3–1.4 (Springer, Berlin, 1980).
- [4] J. P. Keating and M. Sieber, “Calculation of spectral determinants”, *Proc. R. Soc. Lond. A* **447**, 413–437 (1994).
- [5] H. von Koch, “Sur quelques points de la théorie des déterminants infinis”, *Acta Math.* **24**, 89–122 (1901).
- [6] M. Reed and B. Simon, *Methods of Modern Mathematical Physics: Analysis of Operators*, Vol. 4, Chap. XIII.17 (Academic, 1976).
- [7] M. Reed and B. Simon, *Methods of Modern Mathematical Physics: Functional analysis*, Vol. 1, Chap. VI (Academic Press, 1980).
- [8] B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton Univ. Press, 1971).
- [9] B. Simon, “Notes on infinite determinants of Hilbert space operators”, *Adv. Math.* **24**, 244–273 (1977).
- [10] A. Voros, “Spectral functions, special functions and the Selberg zeta function”, *Commun. Math. Phys.* **110**, 439–465 (1987).
- [11] A. Wirzba, “Quantum mechanics and semiclassics of hyperbolic n-disk scattering systems”, *Phys. Rep.* **309**, 1–116 (1999).

Exercises**A40.1. Norm of exponential of an operator.**

Verify inequality (A40.15):

$$\|e^{tA}\| \leq e^{t\|A\|}.$$

Appendix A41

Projects

YOU ARE URGED to work through the essential steps in a project that combines the techniques learned in the course with some application of interest to you for other reasons. It is OK to share computer programs and such, but otherwise each project should be distinct, not a group project. The essential steps are:

- **Dynamics**

1. construct a symbolic dynamics
2. count prime cycles
3. prune inadmissible itineraries, construct transition graphs if appropriate
4. implement a numerical simulator for your problem
5. compute a set of the shortest periodic orbits
6. compute cycle stabilities

- **Averaging, numerical**

1. estimate by numerical simulation some observable quantity, like the escape rate,
2. or check the flow conservation, compute something like the Lyapunov exponent

- **Averaging, periodic orbits**

1. implement the appropriate cycle expansions
2. check flow conservation as function of cycle period truncation, if the system is closed
3. implement desymmetrization, factorization of zeta functions, if dynamics possesses a discrete symmetry

4. compute a quantity like the escape rate as a leading zero of a spectral determinant or a dynamical zeta function.
5. or evaluate a sequence of truncated cycle expansions for averages, such as the Lyapunov exponent or/and diffusion coefficients
6. compute a physically interesting quantity, such as the conductance
7. compute some number of the deterministic and/or quantum eigenvalues, if appropriate