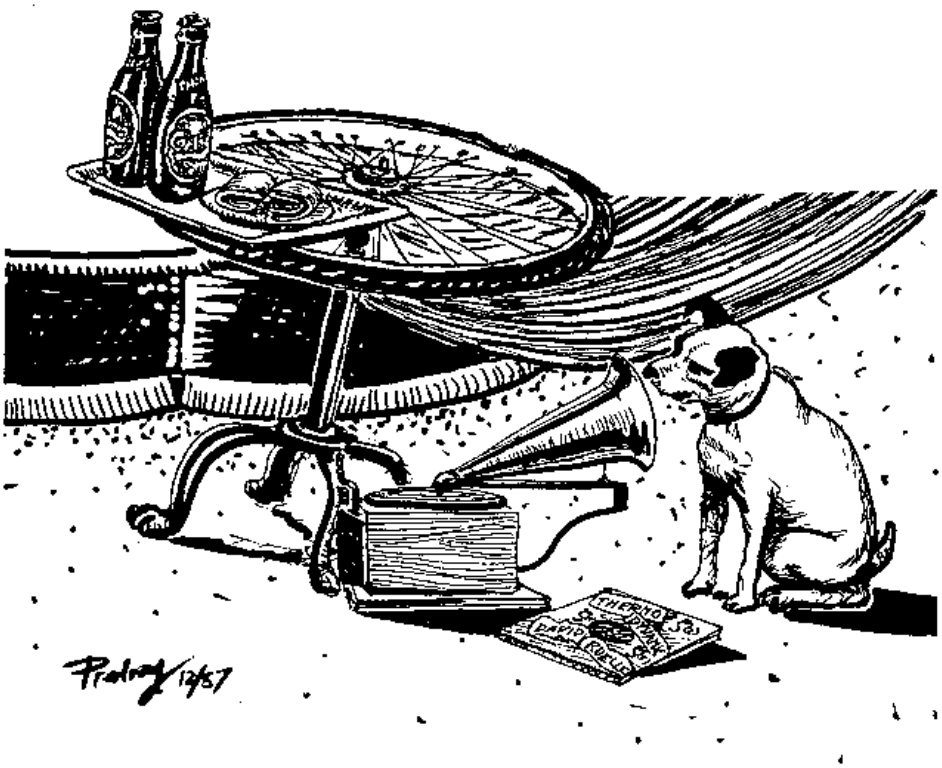
# Chaos: Classical and Quantum

## I: Deterministic Chaos



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

**ChaosBook.org version12, Jun 22 2008** printed June 26, 2008
ChaosBook.org comments to: predrag@nbi.dk

# Contents

## Part I: Classical chaos


Contributors . . . . . iii
Acknowledgements . . . . . vi

**1 Overture** **1**
1.1 Why ChaosBook? . . . . . 2
1.2 Chaos ahead . . . . . 3
1.3 The future as in a mirror . . . . . 4
1.4 A game of pinball . . . . . 9
1.5 Chaos for cyclists . . . . . 13
1.6 Evolution . . . . . 19
1.7 To statistical mechanics . . . . . 21
1.8 What is not in ChaosBook . . . . . 22
résumé 23 commentary 25 guide to exercises 28 exercises 29 references 29

**2 Go with the flow** **32**
2.1 Dynamical systems . . . . . 32
2.2 Flows . . . . . 36
2.3 Computing trajectories . . . . . 40
résumé 41 commentary 41 exercises 43 references 44

**3 Discrete time dynamics** **46**
3.1 Poincaré sections . . . . . 46
3.2 Constructing a Poincaré section . . . . . 53
3.3 Maps . . . . . 54
résumé 57 commentary 57 exercises 59 references 59

**4 Local stability** **61**
4.1 Flows transport neighborhoods . . . . . 61
4.2 Linear flows . . . . . 65
4.3 Stability of flows . . . . . 71
4.4 Neighborhood volume . . . . . 75
4.5 Stability of maps . . . . . 77
résumé 79 commentary 80 exercises 81 references 82

**5 Cycle stability** **83**
5.1 Stability of periodic orbits . . . . . 83
5.2 Cycle Floquet multipliers are cycle invariants . . . . . 87

5.3 Stability of Poincaré map cycles . . . . . 89
5.4 There goes the neighborhood . . . . . 90
résumé 90 commentary 91 exercises 91 references 91

**6 Get straight 93**
6.1 Changing coordinates . . . . . 93
6.2 Rectification of flows . . . . . 95
6.3 Classical dynamics of collinear helium . . . . . 96
6.4 Rectification of maps . . . . . 101
6.5 Rectification of a 1-dimensional periodic orbit . . . . . 102
6.6 Cycle Floquet multipliers are metric invariants . . . . . 103
résumé 104 commentary 104 exercises 106 references 106

**7 Hamiltonian dynamics 108**
7.1 Hamiltonian flows . . . . . 108
7.2 Stability of Hamiltonian flows . . . . . 110
7.3 Symplectic maps . . . . . 113
7.4 Poincaré invariants . . . . . 115
commentary 116 exercises 117 references 118

**8 Billiards 120**
8.1 Billiard dynamics . . . . . 120
8.2 Stability of billiards . . . . . 122
résumé 124 commentary 125 exercises 125 references 126

**9 World in a mirror 128**
9.1 Discrete symmetries . . . . . 129
9.2 Relative periodic orbits . . . . . 136
9.3 Domain for fundamentalists . . . . . 138
9.4 Continuous symmetries . . . . . 140
9.5 Stability . . . . . 143
résumé 144 commentary 145 exercises 147 references 149

**10 Qualitative dynamics, for pedestrians 152**
10.1 Qualitative dynamics . . . . . 152
10.2 Stretch and fold . . . . . 157
10.3 Kneading theory . . . . . 163
10.4 Markov graphs . . . . . 164
10.5 Symbolic dynamics, basic notions . . . . . 166
résumé 170 commentary 170 exercises 171 references 172

**11 Qualitative dynamics, for cyclists 174**
11.1 Recoding, symmetries, tilings . . . . . 175
11.2 Going global: Stable/unstable manifolds . . . . . 178
11.3 Horseshoes . . . . . 179
11.4 Spatial ordering . . . . . 182
11.5 Pruning . . . . . 184
résumé 187 commentary 188 exercises 190 references 191

**12 Fixed points, and how to get them** **195**
12.1 Where are the cycles? . . . . . 196
12.2 One-dimensional mappings . . . . . 200
12.3 Multipoint shooting method . . . . . 201
12.4 Flows . . . . . 204
résumé 207 commentary 207 exercises 209 references 210

**13 Counting** **212**
13.1 How many ways to get there from here? . . . . . 212
13.2 Topological trace formula . . . . . 215
13.3 Determinant of a graph . . . . . 216
13.4 Topological zeta function . . . . . 220
13.5 Counting cycles . . . . . 222
13.6 Infinite partitions . . . . . 226
13.7 Shadowing . . . . . 227
résumé 229 commentary 229 exercises 230 references 233

**14 Transporting densities** **235**
14.1 Measures . . . . . 236
14.2 Perron-Frobenius operator . . . . . 237
14.3 Why not just leave it to a computer? . . . . . 239
14.4 Invariant measures . . . . . 241
14.5 Density evolution for infinitesimal times . . . . . 245
14.6 Liouville operator . . . . . 247
résumé 248 commentary 249 exercises 250 references 252

**15 Averaging** **254**
15.1 Dynamical averaging . . . . . 254
15.2 Evolution operators . . . . . 261
15.3 Lyapunov exponents . . . . . 263
résumé 267 commentary 268 exercises 268 references 269

**16 Trace formulas** **271**
16.1 A trace formula for maps . . . . . 271
16.2 A trace formula for flows . . . . . 275
16.3 An asymptotic trace formula . . . . . 279
résumé 280 commentary 281 exercises 281 references 282

**17 Spectral determinants** **283**
17.1 Spectral determinants for maps . . . . . 283
17.2 Spectral determinant for flows . . . . . 285
17.3 Dynamical zeta functions . . . . . 287
17.4 False zeros . . . . . 290
17.5 Spectral determinants *vs.* dynamical zeta functions . . . . . 291
17.6 All too many eigenvalues? . . . . . 292
résumé 294 commentary 295 exercises 296 references 297

**18 Cycle expansions** **299**
18.1 Pseudocycles and shadowing . . . . . 299
18.2 Construction of cycle expansions . . . . . 302
18.3 Cycle formulas for dynamical averages . . . . . 306
18.4 Cycle expansions for finite alphabets . . . . . 309
18.5 Stability ordering of cycle expansions . . . . . 310
18.6 Dirichlet series . . . . . 313
résumé 314 commentary 315 exercises 317 references 319

**19 Discrete factorization** **320**
19.1 Preview . . . . . 321
19.2 Discrete symmetries . . . . . 323
19.3 Dynamics in the fundamental domain . . . . . 324
19.4 Factorizations of dynamical zeta functions . . . . . 327
19.5 $C_2$ factorization . . . . . 329
19.6 $C_{3v}$ factorization: 3-disk game of pinball . . . . . 331
résumé 332 commentary 333 exercises 334 references 335

**20 Why cycle?** **336**
20.1 Escape rates . . . . . 336
20.2 Natural measure in terms of periodic orbits . . . . . 338
20.3 Flow conservation sum rules . . . . . 340
20.4 Correlation functions . . . . . 341
20.5 Trace formulas *vs.* level sums . . . . . 342
résumé 343 commentary 344 exercises 345 references 346

**21 Why does it work?** **347**
21.1 Linear maps: exact spectra . . . . . 348
21.2 Evolution operator in a matrix representation . . . . . 352
21.3 Classical Fredholm theory . . . . . 355
21.4 Analyticity of spectral determinants . . . . . 357
21.5 Hyperbolic maps . . . . . 361
21.6 The physics of eigenvalues and eigenfunctions . . . . . 363
21.7 Troubles ahead . . . . . 365
résumé 367 commentary 368 exercises 370 references 370

**22 Thermodynamic formalism** **373**
22.1 Rényi entropies . . . . . 373
22.2 Fractal dimensions . . . . . 378
résumé 381 commentary 382 exercises 382 references 383

**23 Intermittency** **385**
23.1 Intermittency everywhere . . . . . 386
23.2 Intermittency for pedestrians . . . . . 388
23.3 Intermittency for cyclists . . . . . 400
23.4 BER zeta functions . . . . . 406
résumé 408 commentary 408 exercises 410 references 411

**24 Deterministic diffusion** **413**
24.1 Diffusion in periodic arrays . . . . . 414
24.2 Diffusion induced by chains of 1-*d* maps . . . . . 418
24.3 Marginal stability and anomalous diffusion . . . . . 424
résumé 428 commentary 430 exercises 432 references 432

**25 Turbulence?** **434**
25.1 Fluttering flame front . . . . . 435
25.2 Infinite-dimensional flows: Numerics . . . . . 438
25.3 Visualization . . . . . 439
25.4 Equilibria of equilibria . . . . . 440
25.5 Why does a flame front flutter? . . . . . 442
25.6 Periodic orbits . . . . . 445
25.7 Intrinsic parametrization . . . . . 445
25.8 Energy budget . . . . . 446
résumé 449 commentary 450 exercises 450 references 451

**26 Noise** **453**
26.1 Deterministic transport . . . . . 453
26.2 Brownian diffusion . . . . . 455
26.3 Weak noise . . . . . 456
26.4 Weak noise approximation . . . . . 458
résumé 459 commentary 459 exercises 461 references 461

**27 Relaxation for cyclists** **464**
27.1 Fictitious time relaxation . . . . . 465
27.2 Discrete iteration relaxation method . . . . . 470
27.3 Least action method . . . . . 474
résumé 474 commentary 475 exercises 478 references 478

**28 Irrationally winding** **480**
28.1 Mode locking . . . . . 481
28.2 Local theory: “Golden mean” renormalization . . . . . 486
28.3 Global theory: Thermodynamic averaging . . . . . 488
28.4 Hausdorff dimension of irrational windings . . . . . 489
28.5 Thermodynamics of Farey tree: Farey model . . . . . 492
résumé 494 commentary 494 exercises 497 references 498

## Part II: Quantum chaos

**29 Prologue** **501**
29.1 Quantum pinball . . . . . 502
29.2 Quantization of helium . . . . . 504
commentary 505 references 506

**30 Quantum mechanics, briefly** **507**
exercises 511

**31 WKB quantization** **512**
31.1 WKB ansatz . . . . . 512
31.2 Method of stationary phase . . . . . 515
31.3 WKB quantization . . . . . 516
31.4 Beyond the quadratic saddle point . . . . . 518
résumé 519 commentary 520 exercises 520 references 521

**32 Semiclassical evolution** **522**
32.1 Hamilton-Jacobi theory . . . . . 522
32.2 Semiclassical propagator . . . . . 530
32.3 Semiclassical Green's function . . . . . 534
résumé 539 commentary 540 exercises 543 references 544

**33 Semiclassical quantization** **545**
33.1 Trace formula . . . . . 545
33.2 Semiclassical spectral determinant . . . . . 550
33.3 One-dof systems . . . . . 552
33.4 Two-dof systems . . . . . 553
résumé 554 commentary 555 exercises 556 references 556

**34 Quantum scattering** **558**
34.1 Density of states . . . . . 558
34.2 Quantum mechanical scattering matrix . . . . . 562
34.3 Krein-Friedel-Lloyd formula . . . . . 563
34.4 Wigner time delay . . . . . 566
commentary 568 exercises 568 references 569

**35 Chaotic multiscattering** **572**
35.1 Quantum mechanical scattering matrix . . . . . 572
35.2 $N$-scatterer spectral determinant . . . . . 576
35.3 Semiclassical 1-disk scattering . . . . . 580
35.4 From quantum cycle to semiclassical cycle . . . . . 586
35.5 Heisenberg uncertainty . . . . . 589
commentary 589

**36 Helium atom** **591**
36.1 Classical dynamics of collinear helium . . . . . 592
36.2 Chaos, symbolic dynamics and periodic orbits . . . . . 593
36.3 Local coordinates, fundamental matrix . . . . . 597

36.4 Getting ready . . . . . 599
36.5 Semiclassical quantization of collinear helium . . . . . 600
résumé 607 commentary 607 exercises 609 references 610

**37 Diffraction distraction** **611**
37.1 Quantum eavesdropping . . . . . 611
37.2 An application . . . . . 617
résumé 622 commentary 622 exercises 624 references 624

**Epilogue** **626**

**Index** **630**

## Part III: Appendices on ChaosBook.org

**A A brief history of chaos** **638**
A.1 Chaos grows up . . . 641
A.2 Chaos with us . . . 642
A.3 Periodic orbit theory . . . 644
A.4 Death of the Old Quantum Theory . . . 646
commentary 649 references 650

**B Linear stability** **652**
B.1 Linear algebra . . . 652
B.2 Eigenvalues and eigenvectors . . . 654
B.3 Stability of Hamiltonian flows . . . 659
B.4 Monodromy matrix for Hamiltonian flows . . . 660
exercises 662

**C Implementing evolution** **663**
C.1 Koopmania . . . 663
C.2 Implementing evolution . . . 665
commentary 668 exercises 668 references 668

**D Symbolic dynamics techniques** **670**
D.1 Topological zeta functions for infinite subshifts . . . 670
D.2 Prime factorization for dynamical itineraries . . . 678

**E Counting itineraries** **682**
E.1 Counting curvatures . . . 682
exercises 683

**F Finding cycles** **684**
F.1 Newton-Raphson method . . . 684
F.2 Hybrid Newton-Raphson / relaxation method . . . 685

**G Transport of vector fields** **688**
G.1 Evolution operator for Lyapunov exponents . . . 688
G.2 Advection of vector fields by chaotic flows . . . 692
commentary 696 exercises 696 references 697

**H Discrete symmetries of dynamics** **699**
H.1 Preliminaries and definitions . . . 699
H.2 Invariants and reducibility . . . 706
H.3 Lattice derivatives . . . 709
H.4 Periodic lattices . . . 712
H.5 Discrete Fourier transforms . . . 714
H.6 $C_{4v}$ factorization . . . 718
H.7 $C_{2v}$ factorization . . . 722
H.8 Hénon map symmetries . . . 724
commentary 725 exercises 725 references 726

**I Convergence of spectral determinants** **727**
I.1 Curvature expansions: geometric picture . . . . . 727
I.2 On importance of pruning . . . . . 730
I.3 Ma-the-matical caveats . . . . . 731
I.4 Estimate of the *n*th cumulant . . . . . 732

**J Infinite dimensional operators** **734**
J.1 Matrix-valued functions . . . . . 734
J.2 Operator norms . . . . . 736
J.3 Trace class and Hilbert-Schmidt class . . . . . 737
J.4 Determinants of trace class operators . . . . . 739
J.5 Von Koch matrices . . . . . 742
J.6 Regularization . . . . . 744
exercises 746 references 746

**K Statistical mechanics recycled** **748**
K.1 The thermodynamic limit . . . . . 748
K.2 Ising models . . . . . 750
K.3 Fisher droplet model . . . . . 753
K.4 Scaling functions . . . . . 759
K.5 Geometrization . . . . . 762
résumé 769 commentary 769 exercises 769 references 770

**L Noise/quantum corrections** **773**
L.1 Periodic orbits as integrable systems . . . . . 773
L.2 The Birkhoff normal form . . . . . 777
L.3 Bohr-Sommerfeld quantization of periodic orbits . . . . . 778
L.4 Quantum calculation of $\hbar$ corrections . . . . . 780
references 786

**S Solutions** **789**

**T Projects** **864**
T.1 Deterministic diffusion, zig-zag map . . . . . 866
references 870
T.2 Deterministic diffusion, sawtooth map . . . . . 872

# 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. Diakonos 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 2008); ChaosBook.org/version12.

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 acknowledgements.

**Roberto Artuso**

14 Transporting densities . . . . . 235
16.2 A trace formula for flows . . . . . 275
20.4 Correlation functions . . . . . 341
23 Intermittency . . . . . 385
24 Deterministic diffusion . . . . . 413

**Ronnie Mainieri**

2 Flows . . . . . 32
3.2 The Poincaré section of a flow . . . . . 53
4 Local stability . . . . . 61
6.1 Understanding flows . . . . . 95
10.1 Temporal ordering: itineraries . . . . . 152
Appendix A: A brief history of chaos . . . . . 638

**Gábor Vattay**

**Gregor Tanner**

23 Intermittency . . . . . 385
Appendix B.4: Jacobians of Hamiltonian flows . . . . . 660

**Arindam Basu**

Rössler flow figures, tables, cycles in chapters 10, 12 and exercise 12.7

**Ofer Biham**

27.1 Cyclists relaxation method . . . . . 465

**Cristel Chandre**

27.1 Cyclists relaxation method . . . . . 465
27.2 Discrete cyclists relaxation methods . . . . . 470

**Freddy Christiansen**

12.2 One-dimensional mappings . . . . . 200
12.3 Multipoint shooting method . . . . . 201

**Per Dahlqvist**

23 Intermittency . . . . . 385
27.3 Orbit length extremization method for billiards . . . . . 474

**Carl P. Dettmann**

18.5 Stability ordering of cycle expansions . . . . . 310

**Fotis K. Diakonos**

27.2 Discrete cyclists relaxation methods . . . . . 470

**G. Bard Ermentrout**

Exercise 5.1

**Mitchell J. Feigenbaum**

Appendix B.3: Symplectic invariance . . . . . 659

**Jonathan Halcrow**

Example 3.5: Sections of Lorenz flow . . . . . 52
Example 4.6: Stability of Lorenz flow equilibria . . . . . 72
Example 4.7: Lorenz flow: Global portrait . . . . . 74
Example 9.2: Desymmetrization of Lorenz flow . . . . . 134
Example 10.5: Lorenz flow: a 1–*d* return map . . . . . 159
Exercises 9.8, 9.9 and figure 2.4

**Kai T. Hansen**

10.2.1 Unimodal map symbolic dynamics . . . . . 158
13.6 Topological zeta function for an infinite partition . . . . . 226
10.3 Kneading theory . . . . . 163
figures throughout the text

**Rainer Klages**

Figure 24.5

**Yueheng Lan**

Solutions 1.1, 2.1, 2.2, 2.3, 2.4, 2.5, 9.2, 9.1, 10.6, 14.1, 14.2, 14.3, 14.5, 14.7, 14.10, 15.1 and figures 1.9, 9.1, 9.6 10.3,

**Bo Li**

Solutions 30.2, 30.1, 31.1

**Joachim Mathiesen**

15.3 Lyapunov exponents . . . . . 263
Rössler flow figures, tables, cycles in sections 15.3, **??** and exercise 12.7

**Yamato Matsuoka**

Figure 11.3

**Rytis Paškauskas**

4.5.1 Stability of Poincaré return maps . . . . . 78
5.3 Stability of Poincaré map cycles . . . . . 89
Exercises 2.8, 3.1, 4.4 and solution 4.1

**Adam Prügel-Bennet**

Solutions 1.2, 2.10, 8.1, 17.1, 18.2 21.3, 27.1,

**Lamberto Rondoni**

14 Transporting densities . . . . . 235
12.1.1 Cycles from long time series . . . . . 197
20.2.1 Unstable periodic orbits are dense . . . . . 339

**Juri Rolf**

Solution 21.3

**Per E. Rosenqvist**

exercises, figures throughout the text

**Hans Henrik Rugh**

21 Why does it work? . . . . . 347

**Peter Schmelcher**

27.2 Discrete cyclists relaxation methods . . . . . 470

**Evangelos Siminos**

Example 3.5: Sections of Lorenz flow . . . . . 52
Example 4.6: Stability of Lorenz flow equilibria . . . . . 72
Example 4.7: Lorenz flow: Global portrait . . . . . 74
Example 9.2: Desymmetrization of Lorenz flow . . . . . 134
Example 10.5: Lorenz flow: a 1–*d* return map . . . . . 159
Exercises 9.8, 9.9

**Gábor Simon**

Rössler flow figures, tables, cycles in chapters 2, 12 and exercise 12.7

**Edward A. Spiegel**

2 Flows . . . . . 32
14 Transporting densities . . . . . 235

**Luz V. Vela-Arevalo**

7.1 Hamiltonian flows . . . . . 108
Exercises 7.1, 7.3, 7.5

## Acknowledgements

> 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 the Weizmann Institute; P. Hemmer of University of Trondheim; The Max-Planck Institut für Mathematik, Bonn; J. Lowenstein of New York University; Edificio Celi, Milano; Fundaçaõ de Faca, Porto Seguro; and Dr. Dj. Cvitanović, Kostrena, for the hospitality during various stages of this work, and the Carlsberg Foundation and Glen P. Robinson for support.

The authors gratefully acknowledge collaborations and/or stimulating discussions with E. Aurell, V. Baladi, B. Brenner, A. de Carvalho, D.J. Driebe, B. Eckhardt, M.J. Feigenbaum, J. Frøjland, 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, Y. Lan, B. Lauritzen, J. Milnor, M. Nordahl, I. Procaccia, J.M. Robbins, P.E. Rosenqvist, D. Ruelle, G. Russberg, M. Sieber, D. Sullivan, N. Søndergaard, T. Tél, C. Tresser, and D. Wintgen.

We thank Dorte Glass for typing parts of the manuscript; B. Lautrup, J.F Gibson and D. Viswanath for comments and corrections to the preliminary versions of this text; the M.A. Porter for lengthening the manuscript by the 2013 definite articles hitherto missing; M.V. Berry for the quotation on page 638; H. Fogedby for the quotation on page 357; J. Greensite for the quotation on page 5; Ya.B. Pesin for the remarks quoted on page 650; M.A. Porter for the quotations on page 19 and page 644; and E.A. Spiegel for quotation on page 1.

F. Haake's heartfelt lament on page 275 was uttered at the end of the first conference presentation of cycle expansions, in 1988. G.P. Morriss advice to students as how to read the introduction to this book, page 4, was offerred during a 2002 graduate course in Dresden. K. Huang's C.N. Yang interview quoted on page 242 is available on `ChaosBook.org/extras`. T.D. Lee remarks on as to who is to blame, page 32 and page 196, as well as M. Shub's helpful technical remark on page 368 came during the Rockefeller University December 2004 "Feigenbaum Fest ."

Who is the 3-legged dog reappearing throughout the book? Long ago, when we were innocent and knew not Borel measurable $\alpha$ to $\Omega$ 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.

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.

# Chapter 1

# Overture

> If I have seen less far than other men it is because I have stood behind giants.
>
> —Edoardo 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 classical 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 cyclist, somewhat advanced 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

indicates 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 A, 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.

## 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, gas pressures.

That is what we will focus on in ChaosBook. The book is a self-contained graduate textbook on classical 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.

## 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 classical 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 exhibits 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 lengths. The insight of the modern dynamical systems theory 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 reminder 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, ☞ on the margin points at the appropriate section. [section 1.4]

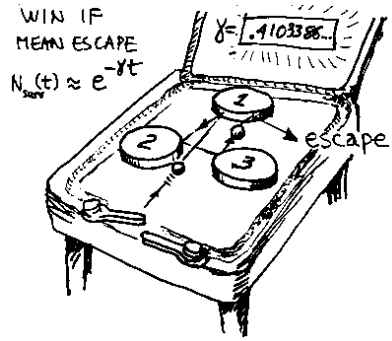



**Figure 1.1:** A physicist's bare bones game of pinball.

## 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 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 [1], 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

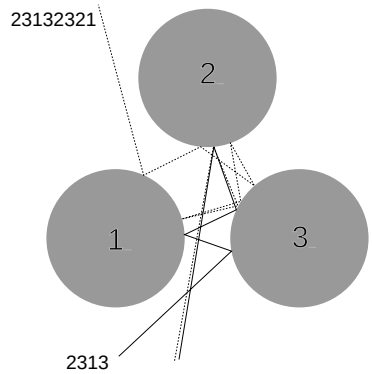



**Figure 1.2:** Sensitivity to initial conditions: two pinballs that start out very close to each other separate exponentially with time.

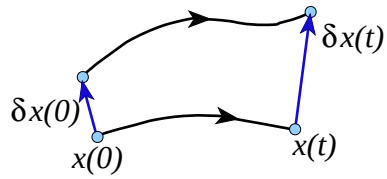



**Figure 1.3:** Unstable trajectories separate with time.

> 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 the classical, 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, in contrast to a stochastic system.

For a stochastic system 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 fool us into regarding it as a stochastic one; 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, and in a finite (and in practice, a very small) number of bounces their separation $\delta\mathbf{x}(t)$ attains the magnitude of $L$, the characteristic linear extent of the whole system, 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 $\lambda$, 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* [section 15.3]

$$T_{\rm Lyap} \approx -\frac{1}{\lambda}\ln|\delta x/L|\,, \tag{1.1}$$

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 13.1]

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

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).

The appellation 'chaos' is a confusing misnomer, as 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,' he means that the system obeys deterministic laws of evolution, but that the outcome is highly sensitive to small uncertainties in the specification of the initial state. The word 'chaos' has in this context taken on a narrow technical meaning. If a deterministic system is locally unstable (positive Lyapunov exponent) and globally mixing (positive entropy)–figure 1.4–it is said to be *chaotic*.

While mathematically correct, the definition of chaos as 'positive Lyapunov + positive entropy' is useless in practice, as a measurement of these quantities is intrinsically asymptotic and beyond reach for systems observed in nature. More

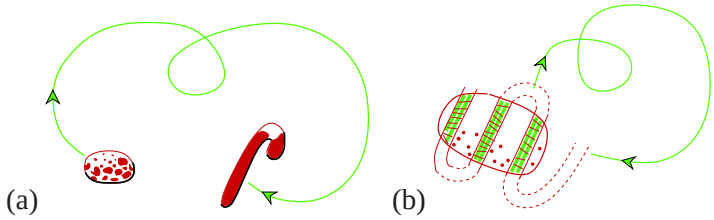
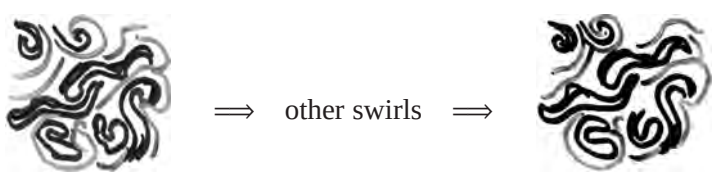



**Figure 1.4:** Dynamics of a *chaotic* dynamical system is (a) everywhere locally unstable (positive Lyapunov exponent) and (b) globally mixing (positive entropy). (A. Johansen)

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 finite time overlap with any other finite region and in this sense spread over the extent of the entire asymptotically accessible state space. Once this is grasped, the focus of theory shifts from attempting to predict individual trajectories (which is impossible) to a description of the geometry of the space of possible outcomes, and evaluation of averages over this space. How this is accomplished is what ChaosBook is about.

A definition of 'turbulence' is even harder to come by. Intuitively, the word refers to irregular behavior of an infinite-dimensional dynamical system 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.

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.

### 1.3.2 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 [4, 5]

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 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.

Even though the experimental evidence for the fractal geometry of nature is circumstantial [2], in studies of probabilistically assembled fractal aggregates we

**Figure 1.5:** Katherine Jones-Smith, 'Untitled 5,' the drawing used by K. Jones-Smith and R.P. Taylor to test the fractal analysis of Pollock's drip paintings [3].

know of nothing better than contemplating such quantities. In deterministic systems we can do *much* better.

## 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 wrote it 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 ☞ 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

[chapter 10]
[chapter 13]

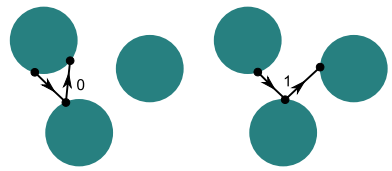



**Figure 1.6:** Binary labeling of the 3-disk pinball trajectories; a bounce in which the trajectory returns to the preceding disk is labeled 0, and a bounce which results in continuation to the third disk is labeled 1.

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.

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.

[chapter 18]

### 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 11]

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

[section 10.5]

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.

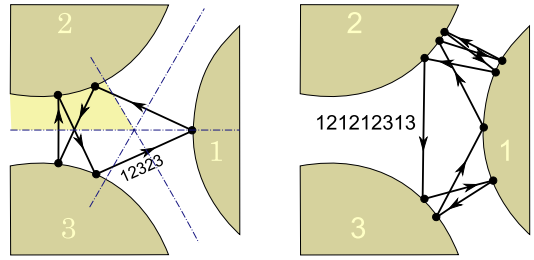



**Figure 1.7:** The 3-disk pinball cycles $\overline{1232}$ and $\overline{121212313}$.

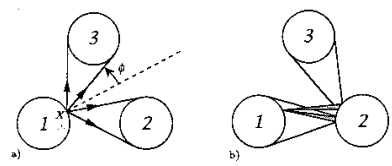



**Figure 1.8:** (a) A trajectory starting out from disk 1 can either hit another disk or escape. (b) Hitting two disks in a sequence requires a much sharper aim, with initial conditions that hit further consecutive disks nested within each other, as in Fig. 1.9.

### 1.4.2 Partitioning with periodic orbits

A trajectory is periodic if it returns to its starting position and momentum. We shall 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 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, see figure 1.9. 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 3.2]

[section 8]

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 figures 1.8 and 1.9. Provided that the disks are sufficiently separated, after $n$ bounces the survivors are divided into $2^n$ distinct strips:

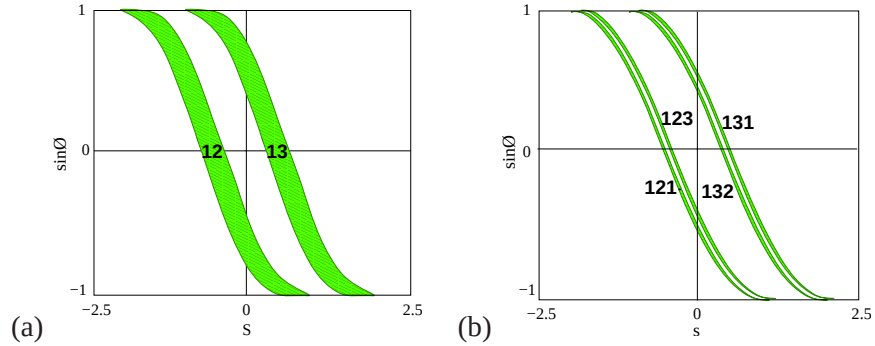



**Figure 1.9:** The 3-disk game of pinball Poincaré section, trajectories emanating from the disk 1 with $x_0 = (s_0, p_0)$ . (a) Strips of initial points $\mathcal{M}_{12}$, $\mathcal{M}_{13}$ which reach disks 2, 3 in one bounce, respectively. (b) Strips of initial points $\mathcal{M}_{121}$, $\mathcal{M}_{131}$ $\mathcal{M}_{132}$ and $\mathcal{M}_{123}$ which reach disks 1, 2, 3 in two bounces, respectively. The Poincaré sections for trajectories originating on the other two disks are obtained by the appropriate relabeling of the strips. Disk radius : center separation ratio a:R = 1:2.5. (Y. Lan)

the $\mathcal{M}_i$th strip consists of all points with itinerary $i = s_1 s_2 s_3 \ldots s_n$, $s = \{1, 2, 3\}$. The unstable cycles as a skeleton of chaos are almost visible here: each such patch contains a periodic point $\overline{s_1 s_2 s_3 \ldots s_n}$ with the basic block infinitely repeated. Periodic 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 $12\ldots$ either eventually escapes or hits the 3rd disk.

### 1.4.3 Escape rate

[example 15.2]

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 classical 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 1.9 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

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

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. $i = 01, 10, 11, \ldots$ 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 20]

$$\hat{\Gamma}_{n+1}/\hat{\Gamma}_n = e^{-\gamma_n} \to e^{-\gamma}. \qquad (1.3)$$

The quantity $\gamma$ is called the *escape rate* from the repeller.

## 1.5 Chaos for cyclists

> Étant données des équations ... et une solution particuliére quelconque de ces équations, on peut toujours trouver une solution périodique (dont la période peut, il est vrai, étre trés 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 ansi dire, la seule bréche par où nous puissions esseyer de pénétrer dans une place jusqu'ici réputée inabordable.
>
> —H. Poincaré, *Les méthodes nouvelles de la méchanique céleste*

We shall now show that the escape rate $\gamma$ 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 $\gamma = 0.41033840776934648933846130781 92\ldots$, you do not need this book. If you have no clue, hang on.

### 1.5.1 How big is my neighborhood?

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.

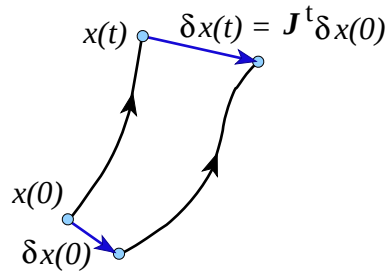



**Figure 1.10:** The fundamental matrix $J^t$ maps an infinitesimal displacement $\delta x$ at $x_0$ into a displacement $J^t(x_0)\delta x$ finite time $t$ later.

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 $x_i(t) + \delta x_i(t)$ is given by the fundamental matrix $J$:

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

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 fundamental matrix is a long exercise - here we just state the result. The fundamental matrix describes the deformation of an infinitesimal neighborhood of $x(t)$ along the flow; its eigenvectors and eigenvalues give the directions and the corresponding rates of expansion or contraction, figure 1.10. 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 maintain their distance along the marginal directions (those whose eigenvalues equal unity in magnitude). [section 8.2]

In our game of pinball the beam of neighboring trajectories is defocused along the unstable eigendirection of the fundamental matrix $M$.

As the heights of the strips in figure 1.9 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 1.9 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,

$$l_i = a_i/|\Lambda_i|\,, \tag{1.4}$$

where $\Lambda_i$ is the unstable eigenvalue of the fundamental matrix $J(x_i)$ 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 14.4]

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 $\Lambda_i$, 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 16.1.1]

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

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 lengths:

$$\Gamma(z) = \sum_{n=1}^{\infty} \Gamma_n z^n \,. \tag{1.5}$$

Recall that for large $n$ the $n$th level sum (1.2) tends to the limit $\Gamma_n \to e^{-n\gamma}$, so the escape rate $\gamma$ is determined by the smallest $z = e^{\gamma}$ for which (1.5) diverges:

$$\Gamma(z) \approx \sum_{n=1}^{\infty} (ze^{-\gamma})^n = \frac{ze^{-\gamma}}{1 - ze^{-\gamma}} \,. \tag{1.6}$$

This is the property of $\Gamma(z)$ that motivated its definition. Next, we devise a formula for (1.5) expressing the escape rate in terms of periodic orbits:

$$\begin{aligned}
\Gamma(z) &= \sum_{n=1}^{\infty} z^n \sum_i^{(n)} |\Lambda_i|^{-1} \\
&= \frac{z}{|\Lambda_0|} + \frac{z}{|\Lambda_1|} + \frac{z^2}{|\Lambda_{00}|} + \frac{z^2}{|\Lambda_{01}|} + \frac{z^2}{|\Lambda_{10}|} + \frac{z^2}{|\Lambda_{11}|} \\
&\quad + \frac{z^3}{|\Lambda_{000}|} + \frac{z^3}{|\Lambda_{001}|} + \frac{z^3}{|\Lambda_{010}|} + \frac{z^3}{|\Lambda_{100}|} + \ldots
\end{aligned} \tag{1.7}$$

For sufficiently small $z$ this sum is convergent. The escape rate $\gamma$ is now given by the leading pole of (1.6), rather than by a numerical extrapolation of a sequence of $\gamma_n$ extracted from (1.3). As any finite truncation $n < n_{\text{trunc}}$ of (1.7) is a polynomial in $z$, convergent for any $z$, finding this pole requires that we know something about $\Gamma_n$ for any $n$, and that might be a tall order.

[section 16.3]

We could now proceed to estimate the location of the leading singularity of $\Gamma(z)$ from finite truncations of (1.7) 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 $\Lambda_p^r$. 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, $p = \overline{0011} = \overline{1001} = \overline{1100} = \overline{0110}$ is prime, but $\overline{0101} = \overline{01}$ 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 length $n_p$ contributes $n_p$ terms to the sum (1.7). Hence (1.7) can be rewritten as

[exercise 13.5]
[section 4.5]

$$\Gamma(z) = \sum_p n_p \sum_{r=1}^{\infty} \left(\frac{z^{n_p}}{|\Lambda_p|}\right)^r = \sum_p \frac{n_p t_p}{1-t_p}, \qquad t_p = \frac{z^{n_p}}{|\Lambda_p|} \tag{1.8}$$

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 $n \le n_p$ 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 (1.8) 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 of the infinite product

$$1/\zeta(z) = \prod_p (1-t_p)\,, \qquad t_p = \frac{z^{n_p}}{|\Lambda_p|}\,. \tag{1.9}$$

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 (1.9). The escape rate is related by (1.6) to a divergence of $\Gamma(z)$, and $\Gamma(z)$ diverges whenever $1/\zeta(z)$ has a zero.

[section 20.1]
[section 17.4]

Easy, you say: "Zeros of (1.9) 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 (1.9) used? We start by computing the lengths 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 length 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 (unless exponentially many more cycles are included). The result of such numerics is a table of the shortest cycles, their periods and their stabilities.

[chapter 12]

[section 27.3]

Now expand the infinite product (1.9), grouping together the terms of the same total symbol string length

$$
\begin{aligned}
1/\zeta &= (1-t_0)(1-t_1)(1-t_{10})(1-t_{100})\cdots \\
&= 1-t_0-t_1-[t_{10}-t_1t_0]-[(t_{100}-t_{10}t_0)+(t_{101}-t_{10}t_1)] \\
&\quad -[(t_{1000}-t_0t_{100})+(t_{1110}-t_1t_{110}) \\
&\quad +(t_{1001}-t_1t_{001}-t_{101}t_0+t_{10}t_0t_1)]-\ldots
\end{aligned}
\tag{1.10}
$$

The virtue of the expansion is that the sum of all terms of the same total length $n$ (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.

[chapter 18]

[section 18.1]

The calculation is now straightforward. We substitute a finite set of the eigenvalues and lengths of the shortest prime cycles into the cycle expansion (1.10), and obtain a polynomial approximation to $1/\zeta$. We then vary $z$ in (1.9) and determine the escape rate $\gamma$ by finding the smallest $z = e^{\gamma}$ for which (1.10) 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 $\overline{0}$, $\overline{1}$ and the 2-cycle $\overline{10}$) 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 18.2.2]

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 1.11; the key observation is that the long orbits are *shadowed* by sequences of shorter orbits.

A typical term in (1.10) is a difference of a long cycle $\{ab\}$ minus its shadowing

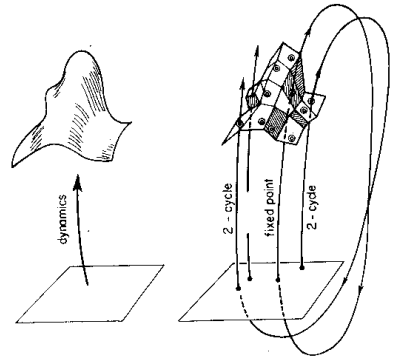



**Figure 1.11:** Approximation to (a) a smooth dynamics by (b) the skeleton of periodic points, together with their linearized neighborhoods. Indicated are segments of two 1-cycles and a 2-cycle that alternates between the neighborhoods of the two 1-cycles, shadowing first one of the two 1-cycles, and then the other.

approximation by shorter cycles $\{a\}$ and $\{b\}$

$$t_{ab} - t_a t_b = t_{ab}(1 - t_a t_b/t_{ab}) = t_{ab}\left(1 - \left|\frac{\Lambda_{ab}}{\Lambda_a \Lambda_b}\right|\right), \tag{1.11}$$

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 figure1.6, 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 $\Lambda_{01}$ 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, figure 1.11.

The orbits that follow the same symbolic dynamics, such as $\{ab\}$ and a 'pseudo orbit' $\{a\}\{b\}$, 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 (1.11) falls off exponentially with the cycle length, and therefore the curvature expansions are expected to be highly convergent.

[chapter 21]

## 1.6 Evolution

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 of 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

$$\Gamma(t) = \frac{\int_{\mathcal{M}} dxdy\, \delta(y - f^t(x))}{\int_{\mathcal{M}} dx} \;\to\; e^{-\gamma t}\,.$$

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

$$\mathcal{L}^t(y, x) = \delta\Big(y - f^t(x)\Big) \tag{1.12}$$

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 a generator of infinitesimal time translations

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

[section 14.6]

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

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 length (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 $n \to \infty$ 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.

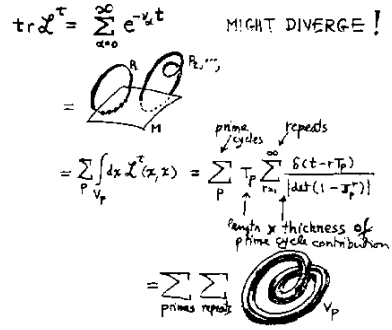



**Figure 1.12:** The trace of an evolution operator is concentrated in tubes around prime cycles, of length $T_p$ and thickness $1/|\Lambda_p|^r$ for the $r$th repetition of the prime cycle $p$.

### 1.6.1 Trace formula

> In physics, when we do not understand something, we give it a name.
>
> —Matthias Neubert

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 (1.12) 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 $\mathrm{tr}\,\mathcal{L}^t$ as a sum over neighborhoods of prime cycles $p$ and their repetitions

[section 16.2]

$$\mathrm{tr}\,\mathcal{L}^t = \sum_p T_p \sum_{r=1}^{\infty} \frac{\delta(t - rT_p)}{\left|\det\left(\mathbf{1} - M_p^r\right)\right|} \ . \tag{1.13}$$

This formula has a simple geometrical interpretation sketched in figure 1.12. After the $r$th return to a Poincaré section, the initial tube $\mathcal{M}_p$ has been stretched out along the expanding eigendirections, with the overlap with the initial volume given by $1/\left|\det\left(\mathbf{1} - M_p^r\right)\right| \to 1/|\Lambda_p|$, the same weight we obtained heuristically in sect. 1.5.1.

The 'spiky' sum (1.13) is disquieting in the way reminiscent of the Poisson resummation formulas of Fourier analysis; the left-hand side is the smooth eigenvalue sum $\mathrm{tr}\, e^{\mathcal{A}t} = \sum e^{s_\alpha t}$, 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 $s_0, s_1, \cdots$ of the classical evolution operator:

[chapter 16]

$$\begin{aligned} \int_{0_+}^{\infty} dt\, e^{-st}\, \mathrm{tr}\,\mathcal{L}^t &= \mathrm{tr}\,\frac{1}{s - \mathcal{A}} = \\ \sum_{\alpha=0}^{\infty} \frac{1}{s - s_\alpha} &= \sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r(\beta\cdot A_p - sT_p)}}{\left|\det\left(\mathbf{1} - M_p^r\right)\right|} \ . \end{aligned} \tag{1.14}$$

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 stability 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

[exercise 4.1]

$$\frac{d}{ds}\ln\det(s-\mathcal{A}) = \mathrm{tr}\,\frac{d}{ds}\ln(s-\mathcal{A}) = \mathrm{tr}\,\frac{1}{s-\mathcal{A}}\,, \tag{1.15}$$

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 17]

$$\det(s-\mathcal{A}) = \exp\left(-\sum_p\sum_{r=1}^{\infty}\frac{1}{r}\frac{e^{-sT_p r}}{\left|\det\left(\mathbf{1}-M_p^r\right)\right|}\right)\,. \tag{1.16}$$

The $1/r$ factor is due to the $s$ integration, leading to the replacement $T_p \to T_p/rT_p$ in the periodic orbit expansion (1.14).

[section 17.5]

The motivation for recasting the eigenvalue problem in this form is sketched in figure 1.13; exponentiation improves analyticity and trades in a divergence of the trace sum for a zero of the spectral determinant. We have now retraced the heuristic derivation of the divergent sum (1.6) and the dynamical zeta function (1.9), but this time with no approximations: formula (1.16) 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.
— Chairman Mao Zedong, a letter to Jiang Qing

The replacement of dynamics of individual trajectories by evolution operators which propagate densities feels like a bit of mathematical voodoo. Actually, something very radical has taken place. 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.

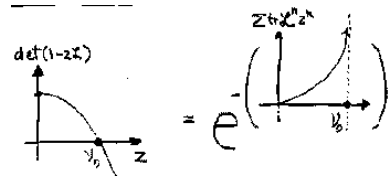



**Figure 1.13:** Spectral determinant is preferable to the trace as it vanishes smoothly at the leading eigenvalue, while the trace formula diverges.

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.

By going to a description in terms of the asymptotic time evolution operators we give up tracking individual trajectories for long times, by trading it in for a very effective 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 (in contrast to the *stosszahlansatz* of Boltzmann). [chapter 24]

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.

The concepts of equilibrium statistical mechanics do help us, however, to understand the ways in which the simple-minded periodic orbit theory falters. A non-hyperbolicity of the dynamics manifests itself in power-law correlations and even 'phase transitions.' [chapter 23]

## 1.8 What is not in ChaosBook

This book offers 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 an honest, hard-working plumber, especially if her hands-on experience is that within the span of even the most careful simulation the dynamics seems to have settled on a (transient?) attractor of dimension less than 3. If rigor, magic, fractals or brains is your thing, read remark 1.4 and beyond.

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

## 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 length. 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 classical 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 (1.4), 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 (1.9) 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 semi-classical quantum mechanics of systems classically chaotic is very much like the classical 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 classical 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** Nonlinear dynamics texts. This text aims to bridge the gap between the physics and mathematics dynamical systems literature. The intended audience is Henri Roux, the perfect physics graduate student with a theoretical bent who does not believe anything he is told. As a complementary presentation we recommend Gaspard's monograph [9] 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 [10], 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 [11] is preferable: an elegant introduction to maps, chaos, period doubling, symbolic dynamics, fractals, dimensions–a good companion to ChaosBook. Introductions more comfortable to physicists is the textbook by Ott [13], 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 first books on nonlinear dynamics. Sprott [14] and Jackson [15] 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. A good introduction to numerical experimentation with physically realistic systems is Tufillaro, Abbott, and Reilly [16]. Korsch and Jodl [17] and Nusse and Yorke [18] also emphasize hands-on approach to dynamics. With this, 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.

**Remark 1.2** ChaosBook based courses. The 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, thermodynamic formalism, period doubling, renormalization operators.

A graduate level introduction to statistical mechanics from the dynamical point view is given by Dorfman [33]; the Gaspard monograph [9] covers the same ground in more depth. Driebe monograph [34] 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?"* [35].

**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 [38] 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.

**Quantum chaos.** Semiclassical propagators, density of states, trace formulas, semiclassical spectral determinants, billiards, semiclassical helium, diffraction, creeping, tunneling, higher-order $\hbar$ corrections. For further reading on this topic, consult the quantum chaos part of ChaosBook.org.

**Remark 1.3** 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 [19, 20], who noted that hidden in the apparent chaos is a rigid skeleton, a tree of *cycles* (periodic orbits) of increasing lengths 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 classical and quantum dynamical systems ever since. We refer the reader to the reprint selection [21] for an overview of some of that literature.

**Remark 1.4** If you seek rigor? If you find ChaosBook not rigorous enough, you should turn to the mathematics literature. The most extensive reference is the treatise by Katok and Hasselblatt [22], an impressive compendium of modern dynamical systems theory. The fundamental papers in this field, all still valuable reading, are Smale [23], Bowen [24] and Sinai [26]. 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 [25]. Markov partitions were introduced by Sinai in ref. [27]. The classical text (though certainly not an easy read) on the subject of dynamical zeta functions is Ruelle's *Statistical Mechanics, Thermodynamic Formalism* [28]. 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 [29]. Further excellent mathematical references on thermodynamic formalism are Parry and Pollicott's monograph [30] with emphasis on the symbolic dynamics aspects of the formalism, and Baladi's clear and compact reviews of the theory of dynamical zeta functions [31, 32].

**Remark 1.5** 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 1.6** Sorry, no shmactals! ChaosBook skirts mathematics and empirical practice of fractal analysis, such as Hausdorff and fractal dimensions. Addison's introduction to fractal dimensions [37] offers a well-motivated entry into this field. While in studies of probabilistically assembled fractals such as Diffusion Limited Aggregates (DLA) better

measures of 'complexity' are lacking, for deterministic systems there are much better, physically motivated and experimentally measurable quantities (escape rates, diffusion coefficients, spectrum of helium, ...) that we focus on here.

**Remark 1.7** 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 ref. [39, 40, 41, 42].

## 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 18.2]

- **Dynamics**
  1. count prime cycles, exercise 1.1, exercise 9.2, exercise 10.1
  2. pinball simulator, exercise 8.1, exercise 12.4
  3. pinball stability, exercise 9.3, exercise 12.4
  4. pinball periodic orbits, exercise 12.5, exercise 12.6
  5. helium integrator, exercise 2.10, exercise 12.8
  6. helium periodic orbits, exercise 12.9
- **Averaging, numerical**
  1. pinball escape rate, exercise 15.3
- **Averaging, periodic orbits**
  1. cycle expansions, exercise 18.1, exercise 18.2
  2. pinball escape rate, exercise 18.4, exercise 18.5
  3. cycle expansions for averages, exercise 18.1, exercise 20.3
  4. cycle expansions for diffusion, exercise 24.1
  5. pruning, Markov graphs, exercise 13.7
  6. desymmetrization exercise 19.1
  7. intermittency, phase transitions, exercise 23.6

The exercises that you should do have **underlined titles**. The rest (**smaller type**) are optional. Difficult problems are marked by any number of *** stars. If you solve one of those, it is probably worth a publication. Solutions to some of the problems are available on `ChaosBook.org`. 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.

## 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. 13.4). Show that the 3-disk pinball has $3 \cdot 2^{n-1}$ itineraries of length $n$. List periodic orbits of lengths 2, 3, 4, 5, $\cdots$. Verify that the shortest 3-disk prime cycles are 12, 13, 23, 123, 132, 1213, 1232, 1323, 12123, $\cdots$. Try to sketch them.

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 18 you should be in position to make very accurate estimates).

## References

[1.1] G.W. Leibniz, *Von dem Verhängnisse*.

[1.2] D. Avnir, O. Biham, D. Lidar and O. Malcai, "Is the Geometry of Nature Fractal?," *Science* **279**, 39 (1998).

[1.3] R. Kennedy, "The Case of Pollock's Fractals Focuses on Physics," *New York Times* (Dec. 2, 2006).

[1.4] R. P. Taylor, A. P. Micolich and D. Jonas, "Fractal analysis of Pollock's drip paintings," *Nature* **399**, 422 (1999).

[1.5] K. Jones-Smith and H. Mathur, "Fractal Analysis: Revisiting Pollock's drip paintings," *Nature* **444**, E9 (2006); R. P. Taylor, A. P. Micolich and D. Jonas, "Fractal Analysis: Revisiting Pollock's drip paintings (Reply)," *Nature* **444**, E10 (2006).

[1.6] T. Li and J. Yorke, "Period 3 implies chaos," *Amer. Math. Monthly* **82**, 985 (1975).

[1.7] P. Cvitanović, B. Eckhardt, P.E. Rosenqvist, G. Russberg and P. Scherer, "Pinball Scattering," in G. Casati and B. Chirikov, eds., *Quantum Chaos* (Cambridge U. Press, Cambridge 1993).

[1.8] K.T. Hansen, *Symbolic Dynamics in Chaotic Systems*, Ph.D. thesis (Univ. of Oslo, 1994);
http://ChaosBook.org/projects/KTHansen/thesis

[1.9] P. Gaspard, *Chaos, Scattering and Statistical Mechanics* (Cambridge U. Press, Cambridge 1998).

[1.10] S.H. Strogatz, *Nonlinear Dynamics and Chaos* (Addison-Wesley 1994).

[1.11] K.T. Alligood, T.D. Sauer and J.A. Yorke, *Chaos, an Introduction to Dynamical Systems* (Springer, New York 1996)

[1.12] T. Tél and M. Gruiz, *Chaotic Dynamics: An Introduction Based on Classical Mechanics* (Cambridge U. Press, Cambridge 2006).

[1.13] E. Ott, *Chaos in Dynamical Systems* (Cambridge U. Press, Cambridge 1993).

[1.14] J. C. Sprott, *Chaos and Time-Series Analysis* (Oxford University Press, Oxford, 2003)

[1.15] E. Atlee Jackson, *Perspectives of nonlinear dynamics* (Cambridge U. Press, Cambridge, 1989).

[1.16] N.B. Tufillaro, T.A. Abbott, and J.P. Reilly, *Experimental Approach to Nonlinear Dynamics and Chaos* (Addison Wesley, Reading MA, 1992).

[1.17] H.J. Korsch and H.-J. Jodl, *Chaos. A Program Collection for the PC*, (Springer, New York 1994).

[1.18] H.E. Nusse and J.A. Yorke, *Dynamics: Numerical Explorations* (Springer, New York 1997).

[1.19] H. Poincaré, *Les méthodes nouvelles de la méchanique céleste* (Guthier-Villars, Paris 1892-99)

[1.20] For a very readable exposition of Poincaré's work and the development of the dynamical systems theory see J. Barrow-Green, *Poincaré and the Three Body Problem*, (Amer. Math. Soc., Providence R.I., 1997), and F. Diacu and P. Holmes, *Celestial Encounters, The Origins of Chaos and Stability* (Princeton Univ. Press, Princeton NJ 1996).

[1.21] R.S. MacKay and J.D. Miess, *Hamiltonian Dynamical Systems* (Adam Hilger, Bristol 1987).

[1.22] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems* (Cambridge U. Press, Cambridge 1995).

[1.23] S. Smale, "Differentiable Dynamical Systems," *Bull. Am. Math. Soc.* **73**, 747 (1967).

[1.24] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms, Springer Lecture Notes in Math.* **470** (1975).

[1.25] D.V. Anosov and Ya.G. Sinai, "Some smooth ergodic systems," *Russ. Math. Surveys* **22**, 103 (1967).

[1.26] Ya.G. Sinai, "Gibbs measures in ergodic theory," *Russ. Math. Surveys* **166**, 21 (1972).

[1.27] Ya.G. Sinai, "Construction of Markov partitions," *Funkts. Analiz i Ego Pril.* **2**, 70 (1968). English translation: *Functional Anal. Appl.* **2**, 245 (1968).

[1.28] D. Ruelle, *Statistical Mechanics, Thermodynamic Formalism*, (Addison-Wesley, Reading MA, 1978).

[1.29] D. Ruelle, "Functional determinants related to dynamical systems and the thermodynamic formalism," (Lezioni Fermiane, Pisa), preprint IHES/P/95/30 (March 1995).

[1.30] W. Parry and M. Pollicott, *Zeta Functions and the Periodic Structure of Hyperbolic Dynamics*, *Astérisque* **187–188** (Société Mathématique de France, Paris 1990).

[1.31] V. Baladi, "Dynamical zeta functions," in B. Branner and P. Hjorth, eds., *Real and Complex Dynamical Systems* (Kluwer, Dordrecht, 1995).

[1.32] V. Baladi, *Positive Transfer Operators and Decay of Correlations* (World Scientific, Singapore 2000).

[1.33] J. R. Dorfman, *An Introduction to Chaos in Nonequilibrium Statistical Mechanics* (Cambridge U. Press, Cambridge 1999).

[1.34] D.J. Driebe, *Fully Chaotic Map and Broken Time Symmetry* (Kluwer, Dordrecht, 1999).

[1.35] J. Bricmont, 'Science of Chaos or Chaos in Science?," in: *The Flight from Science and Reason*, P.R. Gross, N. Levitt, and M.W. Lewis, eds., *Annals of the New York Academy of Sciences* **775**; `mp_arc 96-116.ps.gz`

[1.36] V.I. Arnold, *Mathematical Methods in Classical Mechanics* (Springer-Verlag, Berlin, 1978).

[1.37] P. S. Addison *Fractals and chaos: an illustrated course*, (Inst. of Physics Publishing, Bristol 1997).

[1.38] P. Holmes, J.L. Lumley and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge U. Press, Cambridge 1996).

[1.39] S.J. Schiff, et al. "Controlling chaos in the brain," *Nature* **370**, 615 (1994).

[1.40] F. Moss, "Chaos under control," *Nature* **370**, 596 (1994).

[1.41] J. Glanz, "Do chaos-control techniques offer hope for epilepsy?" *Science* **265**, 1174 (1994).

[1.42] J. Glanz, "Mastering the Nonlinear Brain," *Science* **227**, 1758 (1997).

[1.43] Poul Martin Møller, *En dansk Students Eventyr [The Adventures of a Danish Student]* (Copenhagen 1824).

# Chapter 2

# Go with the flow

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

— T.D. Lee

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

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 coarse brush sketch–a full description of all possible behaviors of dynamical systems is beyond human ken. Anyway, for a novice there is no shortcut through this lengthy detour; a sophisticated traveler might prefer to skip this well-trodden territory and embark upon the journey at chapter14.

fast track:
chapter 14, p. 235

## 2.1 Dynamical systems

[section 1.3]

In a dynamical system we observe the world as a function of time. We express our observations as numbers and record how they change with time; 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. Against the daily motion of the stars from East to West, the planets distinguish themselves by moving among the fixed stars. Ancients discovered that by knowing a sequence of planet's positions–latitudes and longitudes–its future position could be predicted.

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

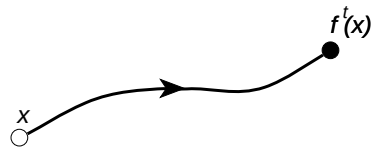



**Figure 2.1:** A trajectory traced out by the evolution rule $f^t$. Starting from the state space point $x$, after a time $t$, the point is at $f^t(x)$.

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* or *phase space* $\mathcal{M}$. As the system changes, so does the *representative point* in state space. We refer to the evolution of such points as *dynamics*, and the function $f^t$ which specifies where the representative point is at time $t$ as the *evolution rule*.

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. The state space can be enlarged, in the hope that in a sufficiently large state space it is possible to determine an evolution rule, so we imagine that knowing the state of the atmosphere, measured over many points over the entire planet should be sufficient to determine the temperature tomorrow. Even that 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. [chapter 11]

Locally, the state space $\mathcal{M}$ looks like $\mathbb{R}^d$, meaning that $d$ numbers are sufficient to determine what will happen next. Globally, it may be a more complicated manifold formed by patching together several pieces of $\mathbb{R}^d$, forming a torus, a cylinder, or some other geometric object. 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 $i = 1, 2, 3, \ldots, d$. The evolution rule $f^t : \mathcal{M} \to \mathcal{M}$ tells us where a point $x$ is in $\mathcal{M}$ after a time interval $t$.

| 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 it is not necessary to know anything else. The time

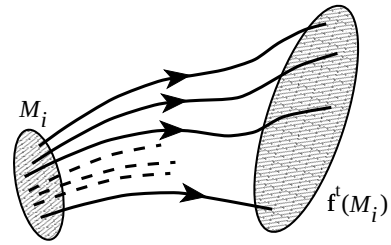



**Figure 2.2:** The evolution rule $f^t$ can be used to map a region $\mathcal{M}_i$ of the state space into the region $f^t(\mathcal{M}_i)$.

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*. Actually, 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 trajectories: given an initial point $x_0$, the evolution rule traces out a sequence of points $x(t) = f^t(x_0)$, the *trajectory* through the point $x_0 = x(0)$. A trajectory is parameterized by the time $t$ and thus belongs to $(f^t(x_0), t) \in \mathcal{M}\times\mathbb{R}$. [exercise 2.1] By extension, we can also talk of the evolution of a region $\mathcal{M}_i$ of the state space: just apply $f^t$ to every point in $\mathcal{M}_i$ to obtain a new region $f^t(\mathcal{M}_i)$, as in figure 2.2.

Because $f^t$ is a single-valued function, any point of the trajectory can be used to label the trajectory. If we mark the trajectory by its initial point $x_0$, we are describing it in the *Lagrangian coordinates*. We can regard the transport of the material point at $t = 0$ to its current point $x(t) = f^t(x_0)$ as a coordinate transformation from the Lagrangian coordinates to the *Eulerian coordinates*.

The subset of points $\mathcal{M}_{x_0} \subset \mathcal{M}$ that belong to the infinite-time trajectory of a given point $x_0$ is called the *orbit* of $x_0$; we shall talk about forward orbits, backward orbits, periodic orbits, etc.. For a flow, an orbit is a smooth continuous curve; for a map, it is a sequence of points. An orbit is a *dynamically invariant* notion. While "trajectory" refers to a state $x(t)$ at time instant $t$, "orbit" refers to the totality of states that can be reached from $x_0$, with state space $\mathcal{M}$ foliated 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).

### 2.1.1 A classification of possible motions?

What are the possible trajectories? This is a grand question, and there are many answers, chapters to follow offering some. Here is the first attempt to classify all possible trajectories:

| | | |
|---|---|---|
| stationary: | $f^t(x) = x$ | for all $t$ |
| periodic: | $f^t(x) = f^{t+T_p}(x)$ | for a given minimum period $T_p$ |
| aperiodic: | $f^t(x) \neq f^{t'}(x)$ | for all $t \neq t'$ . |

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. Periodic orbits form a

very small 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 notion of periodic motion generalized to quasiperiodic (superposition of $M$ incommesurate frequencies) motion on a smooth torus, and families of solutions related by a continuous symmetry.

The ancients tried to make sense of all dynamics in terms of periodic motions; epicycles, integrable systems. The embarassing truth is that for a generic dynamical systems 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 trajectory never returns

$$f^t(x) \notin \mathcal{M}_0 \qquad \text{for all} \quad t > t_{\rm min}\,. \tag{2.1}$$

In physics literature, the dynamics of such 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 trajectories, 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

$$f^t(x) \in \mathcal{M}_0\,. \tag{2.2}$$

In other words, the trajectory of a non-wandering point reenters the neighborhood $\mathcal{M}_0$ infinitely often. We shall denote by $\Omega$ the *non–wandering set* of $f$, i.e., the union of all the non-wandering points of $\mathcal{M}$. The set $\Omega$, the non–wandering set of $f$, is the 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 foward evolution. The attractor can be a fixed point, a periodic orbit, 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. [example 2.3]

Conversely, if we can enclose the non–wandering set $\Omega$ by a connected state space volume $\mathcal{M}_0$ and then show that almost all points within $\mathcal{M}_0$, but not in $\Omega$, eventually exit $\mathcal{M}_0$, we refer to the non–wandering set $\Omega$ as a *repeller*. An example of a repeller is not hard to come by–the pinball game of sect. 1.3 is a simple chaotic repeller.

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 truth. As longer and longer cycles approximate more and more accurately finite segments of aperiodic trajectories, we shall 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

There is no beauty without some strangeness.
—William Blake

A *flow* is a continuous-time dynamical system. The evolution rule $f^t$ is a family of mappings of $\mathcal{M} \to \mathcal{M}$ 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 '$\circ$ :' [appendix H.1]

$$f^{t+s} = f^t \circ f^s = f^t(f^s)\,. \tag{2.3}$$

The family of mappings $f^t(x)$ thus forms a continuous (forward 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. In exceedingly many situations of interest–for times beyond the Lyapunov time, for

asymptotic attractors, for dissipative partial differential equations, for systems with noise, for non-invertible maps–the dynamics cannot be run backwards in time, hence, the circumspect emphasis on *semi*groups. 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 as

$$x(t+\tau) = f^{t+\tau}(x_0) = f(f(x_0,t),\tau) \tag{2.4}$$

and express the time derivative of a trajectory at point $x(t)$,

[exercise 2.3]

$$\left.\frac{dx}{d\tau}\right|_{\tau=0} = \partial_\tau f(f(x_0,t),\tau)|_{\tau=0} = \dot{x}(t)\,. \tag{2.5}$$

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

$$v(x) = \dot{x}(t)\,. \tag{2.6}$$

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 (2.6), as they are 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 (2.6). At each point of the state space a vector indicates the local direction in which the trajectory evolves. The length of the vector $|v(x)|$ is proportional to the speed at the point $x$, and the direction and length of $v(x)$ changes from point to point. When the state space is a complicated manifold embedded in $\mathbb{R}^d$, one can no longer think of the vector field as being embedded in the state space. Instead, we have to imagine that each point $x$ of state space has a different tangent plane $T\mathcal{M}_x$ attached to it. The vector field lives in the union of all these tangent planes, a space called the *tangent bundle* $\mathbf{T}\mathcal{M}$.

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

$$\begin{aligned} \dot{x}(t) &= y(t) \\ \dot{y}(t) &= -0.15\,y(t) + x(t) - x(t)^3 \end{aligned} \tag{2.7}$$

*plotted in figure 2.3. The velocity vectors are drawn superimposed over the configuration coordinates* $(x(t), y(t))$ *of state space* $\mathcal{M}$*, but they belong to a different space, the* tangent bundle $\mathbf{T}\mathcal{M}$.

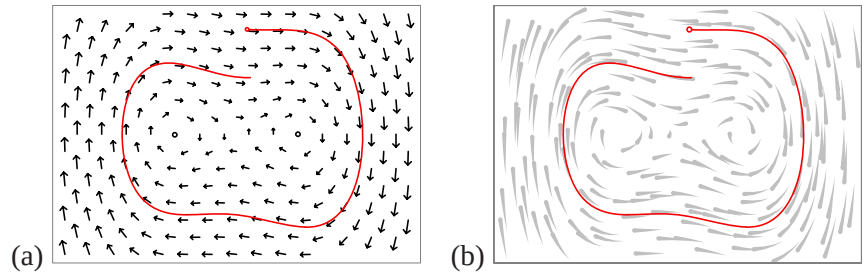



**Figure 2.3:** (a) The 2-dimensional vector field for the Duffing system (2.7), together with a short trajectory segment. (b) The flow lines. Each 'comet' represents the same time interval of a trajectory, starting at the tail and ending at the head. The longer the comet, the faster the flow in that region.

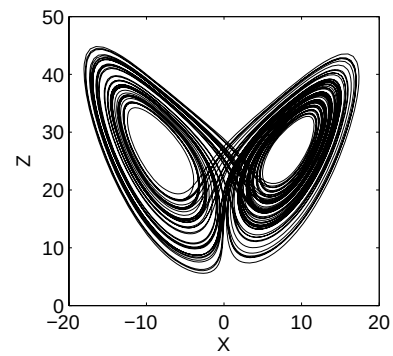



**Figure 2.4:** Lorenz "butterfly" strange attractor. (J. Halcrow)

$$\text{If} \qquad v(x_q) = 0\,, \tag{2.8}$$

$x_q$ is an *equilibrium point* (also referred to as a *stationary*, *fixed*, *critical*, *invariant*, *rest*, *stagnation* point, *zero* of the vector field $v$, or *steady state* - our usage is 'equilibrium' for a flow, 'fixed point' for a map), and the trajectory remains forever stuck at $x_q$. Otherwise the trajectory passing through $x_0$ at time $t = 0$ can be obtained by integrating the equations (2.6):

$$x(t) = f^t(x_0) = x_0 + \int_0^t d\tau\, v(x(\tau))\,, \qquad x(0) = x_0\,. \tag{2.9}$$

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

$$\frac{dy}{d\tau} = w(y,\tau)\,, \tag{2.10}$$

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 [exercise 2.4] [exercise 2.5]

$$v(x) = \begin{bmatrix} w(y,\tau) \\ 1 \end{bmatrix}. \tag{2.11}$$

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

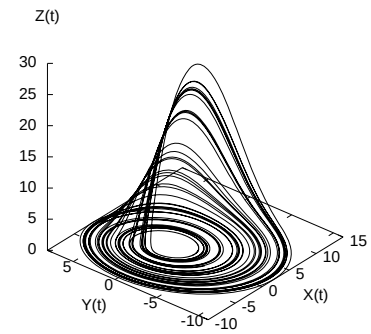



**Figure 2.5:** A trajectory of the Rössler flow at time $t = 250$. (G. Simon)

**Example 2.2** ***Lorenz strange attractor:*** *Edward Lorenz arrived at the equation*

$$\dot{x} = v(x) = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} \sigma(y-x) \\ \rho x - y - xz \\ xy - bz \end{bmatrix} \tag{2.12}$$

*by a drastic simplification of the Rayleigh-Benard flow. Lorenz fixed $\sigma = 10$, $b = 8/3$, and varied the "Rayleigh number" $\rho$. 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 at*

[remark 2.2]

$$x_{EQ1,2} = (\pm\sqrt{b(\rho-1)}, \pm\sqrt{b(\rho-1)}, \rho-1)\,, \tag{2.13}$$

*We shall not explore the Lorenz flow dependence on the $\rho$ parameter in what follows, but here is a brief synopsis: the $EQ_0$ $1d$ unstable manifold closes into a homoclinic orbit at $\rho = 13.56\ldots$. Beyond that, an infinity of associated periodic orbits are generated, until $\rho = 24.74\ldots$, 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 2.4. (Continued in example 3.5.)*

**Example 2.3** ***The Rössler flow–A flow with a strange attractor:*** *The Duffing flow of figure 2.3 is bit of a bore–every trajectory 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*

$$\dot{x} = -y\,, \qquad \dot{y} = x\,. \tag{2.14}$$

*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*

$$\dot{x} = -y\,, \qquad \dot{y} = x + ay\,, \qquad a > 0\,, \tag{2.15}$$

*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 $\infty$, kick it into 3rd dimension when $x$ reaches some value $c$ by adding*

$$\dot{z} = b + z(x-c)\,, \qquad c > 0\,. \tag{2.16}$$

*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 *(for definitiveness, we fix the parameters $a$, $b$, $c$ in what follows):*

$$\begin{aligned} \dot{x} &= -y - z \\ \dot{y} &= x + ay \\ \dot{z} &= b + z(x-c)\,, \qquad a = b = 0.2\,, \quad c = 5.7\,. \end{aligned} \tag{2.17}$$

*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 (2.8):*

$$\begin{aligned} x_\pm &= \frac{c \pm \sqrt{c^2 - 4ab}}{2a}(a, -1, 1) \\ (x_-, y_-, z_-) &= (\,0.0070,\ -0.0351,\ 0.0351\,) \\ (x_+, y_+, z_+) &= (\,5.6929,\ -28.464,\ 28.464\,) \end{aligned} \tag{2.18}$$

*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 other solutions look like we need to resort to numerical integration. A typical numerically integrated long-time trajectory is sketched in figure 2.5. As we shall show in sect. 4.1, for this flow any finite volume of initial conditions shrinks with time, so the flow is* contracting*. 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 2.5 and similar figures in what follows are examples of 'strange attractors.' (continued in exercise 2.8 and example 3.4)* [exercise 3.5]

*(R. Paškauskas)*

fast track:
chapter 3, p. 46

## 2.3 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 (2.6). The simplest is the Euler integrator which advances the trajectory by $\delta\tau \times$ velocity at each time step:

$$x_i \to x_i + v_i(x)\delta\tau \,. \tag{2.19}$$

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 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 2.10]

## Résumé

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, and 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.

## Commentary

**Remark 2.1** Rössler and Duffing flows. The Duffing system (2.7) arises in the study of electronic circuits [2]. The Rössler flow (2.17) is the simplest flow which exhibits many of the key aspects of chaotic dynamics. We shall us the Rössler and the 3-pinball (see chapter 8) systems throughout ChaosBook to motivate the notions of Poincaré sections, return maps, symbolic dynamics, cyce expansions, etc., etc.. The Rössler flow was introduced in ref. [3] as a set of equations describing no particular physical system, but capturing the essence of chaos in a simplest imaginable smooth flow. 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. [5, 8, 11]. We recommend in particular the inimitable Abraham and Shaw illustrated classic [6] for its beautiful sketches of the Rössler and many other flows. Timothy Jones [19] has a number of interesting simulations on a Drexel website.

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 Poincaré map computed in exercise 3.2. Its volume contraction rate is computed in exercise 4.3, its topology investigated in exercise 4.4, and its Lyapunov exponents evaluated in exercise 15.4. The shortest Rössler flow cycles are computed and tabulated in exercise 12.7.

**Remark 2.2** Lorenz equation. The Lorenz equation (2.12) is the most celebrated early illustration of "deterministic chaos" [13] (but not the first - the honor goes to Dame Cartwright [27]). Lorenz's paper, which can be found in reprint collections refs. [14, 15], is a pleasure to read, and is still one of the best introductions to the physics motivating such models. For a geophysics derivation, see Rothman course notes [7]. The equations, a set of ODEs in $\mathbb{R}^3$, exhibit strange attractors [28, 29, 30]. Frøyland [16] has a nice brief discussion of Lorenz flow. Frøyland and Alfsen [17] plot many periodic and heteroclinic orbits of the Lorenz flow; some of the symmetric ones are included in ref. [16]. Guckenheimer-Williams [18] and Afraimovich-Bykov-Shilnikov [19] offer in-depth discussion of the Lorenz equation. The most detailed study of the Lorenz equation was undertaken by Sparrow [21]. For a physical interpretation of $\rho$ as "Rayleigh number." see Jackson [24] and Seydel [25]. Lorenz truncation to 3 modes is so drastic that the model bears no relation to the physical hydrodynamics problem that motivated it. For a detailed pictures of Lorenz invariant manifolds consult Vol II of Jackson [24]. Lorenz attractor is a very thin fractal – as we saw, stable manifold thinckness is of order $10^{-4}$ – but its fractal structure has been accurately resolved by D. Viswanath [9, 10]. (Continued in remark 9.1.)

**Remark 2.3** Diagnosing chaos. In sect. 1.3.1 we have stated that a deterministic system exhibits 'chaos' if its dynamics is locally unstable (positive Lyapunov exponent) and globally mixing (positive entropy). In sect. 15.3 we shall 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 [15] 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. [16].

**Remark 2.4** Dynamical systems software: J.D. Meiss [13] 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, and a up-to-date software list [14], with links to DSTool, xpp, AUTO, etc.. Springer on-line *Encyclopaedia of Mathematics* maintains links to dynamical systems software packages on eom.springer.de/D/d130210.htm.

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. **Trajectories do not intersect.** A trajectory in the state space $\mathcal{M}$ is the set of points one gets by evolving $x \in \mathcal{M}$ forwards and backwards in time:

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

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 (2.3)

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

Show that it is a commutative semigroup.

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

2.3. **Almost ODE's.**

(a) Consider the point $x$ on $\mathbb{R}$ evolving according $\dot{x} = e^{\dot{x}}$. 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' $\phi$

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

where $x \in \mathbb{R}^d$, and $\phi$ 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 $\phi$.

(b) Show that all extrema of $\phi$ 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. [12]) 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 (2.17). 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 (2.17). How many are there?

(b) Assume that $b = a$. As we shall see, some surprisingly large, and surprisingly small numbers arise in this system. In order to understand their size, introduce parameters

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

Express all the equilibria in terms of $(c, \epsilon, D, p^{\pm})$. Expand equilibria to the first order in $\epsilon$. Note that it makes sense because for $a = b = 0.2$, $c = 5.7$ in (2.17), $\epsilon \approx 0.03$. (continued as 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 were 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 $\ddot{x} = -x$ with initial conditions $x(0) = 2$ and $\dot{x} = 0$ has as solution $x(t) = e^{-t}(1+e^{2t})$. Can your integrator reproduce this solution for the interval $t \in [0, 10]$? Check you solution by plotting the error as compared to the exact result.

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

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

which can be checked–numerically–to be chaotic. As 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 $\approx 2.4$).

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

2.10. **Classical collinear helium dynamics.** In order to apply periodic orbit theory to quantization of helium we shall need to compute classical periodic orbits of the helium system. In this exercise we commence their evaluation for the collinear helium atom (7.6)

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

The nuclear charge for helium is $Z = 2$. Colinear helium has only 3 degrees of freedom and the dynamics can be visualized as a motion in the $(r_1, r_2)$, $r_i \geq 0$ quadrant. In $(r_1, r_2)$-coordinates the potential is singular for $r_i \to 0$ nucleus-electron collisions. These 2-body collisions can be regularized by rescaling the coordinates, with details given in sect. 6.3. In the transformed coordinates $(x_1, x_2, p_1, p_2)$ the Hamiltonian equations of motion take the form

$$\begin{aligned}
\dot{P}_1 &= 2Q_1\left[2 - \frac{P_2^2}{8} - Q_2^2(1 + \frac{Q_2^2}{R^4})\right] \\
\dot{P}_2 &= 2Q_2\left[2 - \frac{P_1^2}{8} - Q_1^2(1 + \frac{Q_1^2}{R^4})\right] \\
\dot{Q}_1 &= \frac{1}{4}P_1Q_2^2\,, \quad \dot{Q}_2 = \frac{1}{4}P_2Q_1^2\,.
\end{aligned} \tag{2.20}$$

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

(a) 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-$d$ state space orbit is by projecting it onto the 2-dimensional $(r_1(t), r_2(t))$ plane. (continued as exercise 3.4)

(Gregor Tanner, Per Rosenqvist)

## References


[2.1] E.N. Lorenz, "Deterministic nonperiodic flow," *J. Atmos. Sci.* **20**, 130 (1963).

[2.2] G. Duffing, *Erzwungene Schwingungen bei veränderlicher Eigenfrequenz* (Vieweg. Braunschweig 1918).

[2.3] O. Rössler, *Phys. Lett.* **57A**, 397 (1976).

[2.4] "Rössler attractor," en.wikipedia.org/wiki/Rossler_map.

[2.5] J. Peinke, J. Parisi, O.E. Rössler, and R. Stoop, *Encounter with Chaos. Self-Organized Hierarchical Complexity in Semiconductor Experiments* (Springer, Berlin 1992).

[2.6] R.H. Abraham, C.D. Shaw, *Dynamics–The Geometry of Behavior* (Addison-Wesley, Redwood, Ca, 1992).

[2.7] D. Rothman, Nonlinear Dynamics I: Chaos (MIT OpenCourseWare 2006).

[2.8] R. Gilmore and M. Lefranc, *The Topology of Chaos* (Wiley, New York, 2002).

[2.9] D. Viswanath, "Symbolic dynamics and periodic orbits of the Lorenz attractor," *Nonlinearity* **16**, 1035 (2003).

[2.10] D. Viswanath, "The fractal property of the Lorenz attractor," *Physica D* **190**, 115 (2004).

[2.11] J.M.T. Thompson and H.B. Stewart *Nonlinear Dynamics and Chaos* (Wiley, New York, 2002).

[2.12] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes* (Cambridge University Press, 1986).

[2.13] J.D. Meiss, *Sci.nonlinear FAQ, Computational Resources*, amath.colorado.edu/faculty/jdm/faq.html.

[2.14] DSWeb Dynamical Systems Software, `www.dynamicalsystems.org`.

[2.15] J. Laskar, *Icarus* **88**, 257 (1990).

[2.16] Ch. Skokos, "Alignment indices: a new, simple method for determining the ordered or chaotic nature of orbits," *J. Phys* **A 34**, 10029 (2001).

[2.17] P. Cvitanović, "Periodic orbits as the skeleton of classical and quantum chaos," *Physica* **D 51**, 138 (1991).

[2.18] M.W. Hirsch, "The dynamical systems approach to differential equations," *Bull. Amer. Math. Soc.* **11**, 1 (1984)

[2.19] T. Jones, *Symmetry of Chaos Animations*, lagrange.physics.drexel.edu/flash.

# Chapter 3

# Discrete time dynamics

Do it again!
—Isabelle, age 3

(R. Mainieri and P. Cvitanović)

THE TIME PARAMETER in the sect. 2.1 definition of a dynamical system can be either continuous or discrete. Discrete time dynamical systems arise naturally from flows; one can observe the flow at fixed time intervals (by strobing it), or one can record the coordinates of the flow when a special event happens (the Poincaré section method). This triggering event can be as simple as vanishing of one of the coordinates, or as complicated as the flow cutting through a curved hypersurface.

## 3.1 Poincaré sections

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

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

Here the *first return function* $\tau(x)$–sometimes referred to as the *ceiling function*–is the time of flight to the next section for a trajectory starting at $x$. 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. In practice one often needs only a local section–a finite hypersurface of codimension 1 volume intersected by a ray of trajectories near to the trajectory of interest. The hypersurface can be specified implicitly through a function $U(x)$ that is zero whenever a point $x$ is on the Poincaré section,

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

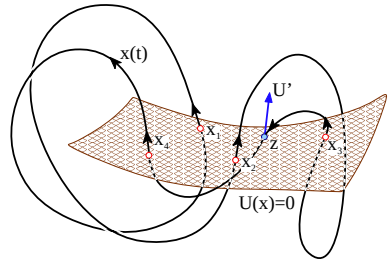



**Figure 3.1:** A $x(t)$ trajectory that intersects a Poincaré section $\mathcal{P}$ at times $t_1, t_2, t_3, t_4$, and closes a cycle $(x_1, x_2, x_3, x_4)$, $x_k = x(t_k) \in \mathcal{P}$ of topological length 4 with respect to this section. Note that the intersections are not normal to the section, and that the crossing $z$ does not count, as it in the wrong direction.

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

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

Second, the gradient $\partial_j U$ 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 Poincaré return map $P(x)$ needs to be supplemented with the orientation condition

$$\begin{aligned} x_{n+1} = P(x_n)\,, \qquad & U(x_{n+1}) = U(x_n) = 0\,, \quad n \in \mathbb{Z}^+ \\ & \sum_{j=1}^{d} v_j(x_n)\partial_j U(x_n) \;>\; 0\,. \end{aligned} \tag{3.4}$$

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

[chapter 15]

With a sufficiently clever choice of a Poincaré section or a set of sections, any orbit of interest intersects a section. 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 $\tau(x_n)$, with the accumulated flight time given by

$$t_{n+1} = t_n + \tau(x_n)\,, \qquad t_0 = 0\,, \qquad x_n \in \mathcal{P}\,. \tag{3.5}$$

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.

A few examples may help visualize this.

**Example 3.1** ***Hyperplane*** $\mathcal{P}$***:*** *The simplest choice of a Poincaré section is a plane* $\mathcal{P}$ *specified by a point (located at the tip of the vector* $r_0$*) and a direction vector* $a$ *perpendicular to the plane. A point* $x$ *is in this plane if it satisfies the condition*

$$U(x) = (x - r_0) \cdot a = 0\,. \tag{3.6}$$

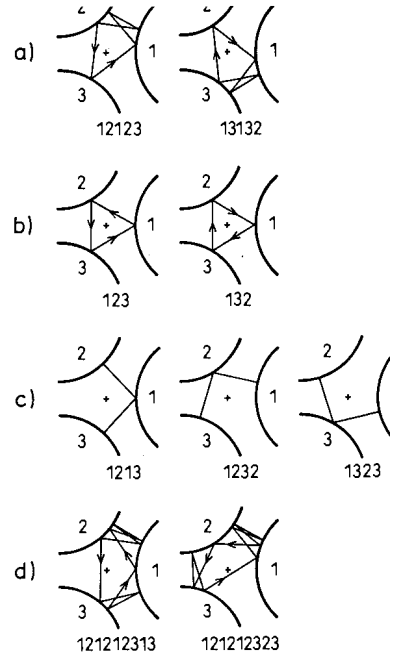



**Figure 3.2:** Some examples of 3-disk cycles: (a) $\overline{12123}$ and $\overline{13132}$ are mapped into each other by the flip across 1 axis. Similarly (b) $\overline{123}$ and $\overline{132}$ are related by flips, and (c) $\overline{1213}$, $\overline{1232}$ and $\overline{1323}$ by rotations. (d) The cycles $\overline{121212313}$ and $\overline{121212323}$ are related by rotation *and* time reversal. These symmetries are discussed in chapter 9. (From ref. [7])

*Consider a circular periodic orbit centered at $r_0$, but not lying in $\mathcal{P}$. It pierces the hyperplane twice; the $(v \cdot a) > 0$ traversal orientation condition (3.4) ensures that the first return time is the full period of the cycle.*

The above flat hyperplane is an *ad hoc* construct; one Poincaré section rarely suffices to capture all of the dynamics of interest. A more insightful picture of the dynamics is obtained by partitioning the state space into $N$ qualitatively distinct regions $\{\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_N\}$ and constructing a Poincaré section $\mathcal{P}_s$ per region. The $d$-dimensional flow is thus reduced reduced to composition [section 10.1]

$$P_{s_n \leftarrow s_{n-1}} \circ \cdots \circ P_{s_2 \leftarrow s_1} \circ P_{s_1 \leftarrow s_0}$$

of a set of $(d-1)$-dimensional maps

$$P_{s_{n+1} \leftarrow s_n} : \quad x_n \mapsto x_{n+1} , \qquad s \in \{1, 2, \ldots, N\} \tag{3.7}$$

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.

A *return map* $P_{s_0}$ from section $\mathcal{P}_{s_0}$ to itself now has a contribution from any admissible (i.e., there exist trajectories that traverse regions $\mathcal{M}_{s_0} \to \mathcal{M}_{s_1} \to \cdots \to \mathcal{M}_{s_n} \to \mathcal{M}_{s_0}$ in the same temporal sequence) periodic sequence of compositions

$$P_{s_0 s_1 \cdots s_{n-1}} = P_{s_0 \leftarrow s_{n-1}} \circ \cdots \circ P_{s_2 \leftarrow s_1} \circ P_{s_1 \leftarrow s_0} \tag{3.8}$$

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 a covering symbolic dynamics, a subject that will will return to in chapter 10. [chapter 10]

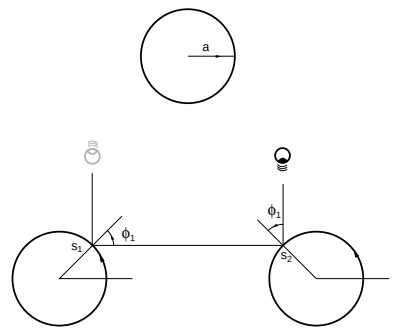



**Figure 3.3:** Poincaré section coordinates for the 3-disk game of pinball.

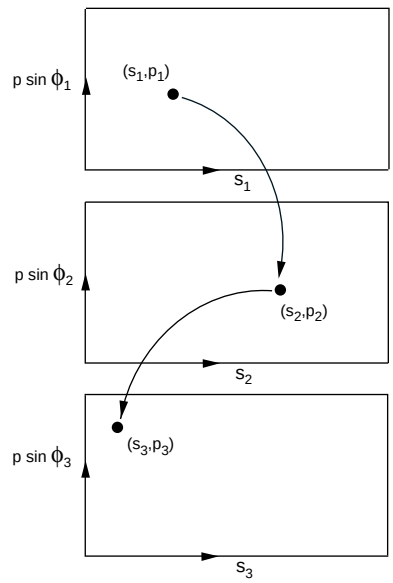



**Figure 3.4:** Collision sequence $(s_1, p_1) \mapsto (s_2, p_2) \mapsto (s_3, p_3)$ from the boundary of a disk to the boundary of the next disk is coded by the Poincaré sections maps sequence $P_{3\leftarrow 2}P_{2\leftarrow 1}$.

**Example 3.2** ***Pinball game, Poincaré dissected.*** *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 3.2 (d), and 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 a 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 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 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 momentum component parallel to the billiard wall at the point of impact, figure 3.3. 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 without further ado. 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*

$$P_{\sigma_{n+1}\leftarrow\sigma_n}: \quad (s_n, p_n) \mapsto (s_{n+1}, p_{n+1}), \qquad \sigma \in \{1, 2, 3\} \tag{3.9}$$

*from the boundary of the disk $j$ to the boundary of the next disk $k$, figure 3.4. The explicit form of this map is easily written down, see sect. 8, but much more economical*

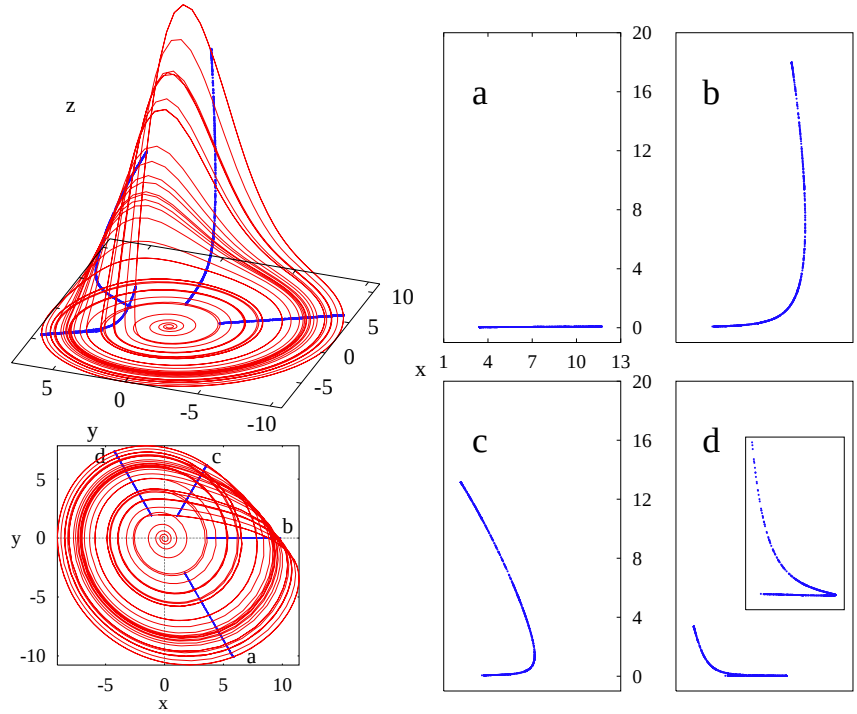



**Figure 3.5:** (Right:) a sequence of Poincaré sections of the Rössler strange attractor, defined by planes through the $z$ axis, oriented at angles (a) $-60^o$ (b) $0^o$, (c) $60^o$, (d) $120^o$, in the $x$-$y$ plane. (Left:) side and $x$-$y$ plane view of a typical trajectory with Poincaré sections superimposed. (R. Paškauskas)

*is the symmetry quotiented version of chapter 9 which replaces the above 6 maps by a* return map *pair* $P_0, P_1$. [chapter 9]

[chapter 8]

*Embedded within* $\mathcal{P}_{12}$, $\mathcal{P}_{13}$ *are four strips* $\mathcal{P}_{121}$, $\mathcal{P}_{123}$, $\mathcal{P}_{131}$, $\mathcal{P}_{132}$ *of initial conditions that survive two bounces, and so forth. Provided that the disks are sufficiently separated, after* $n$ *bounces the survivors are labeled by* $2^n$ *distinct itineraries* $\sigma_1\sigma_2\sigma_3\ldots\sigma_n$.

Billiard dynamics is exceptionally simple - free flight segments, followed by specular reflections at boundaries, thus billiard boundaries are the obvious choice as Poincaré sections. What about smooth, continuous time flows, with no obvious surfaces that would fix the choice of Poincaré sections?

**Example 3.3** ***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, in the pendulum all orbits are loops, so any trajectory will periodically intersect the line, that is the Poincaré section, at one point.*

*Consider next a pendulum with friction, such as the unforced Duffing system plotted in figure 2.3. 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 of the equilibrium points; the Duffing oscillator at rest.*

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

**Example 3.4** ***Rössler flow:*** *Consider figure 2.5, a typical trajectory of the 3-dimensional Rössler flow (2.17). It wraps around the* $z$ *axis, so a good choice for a*

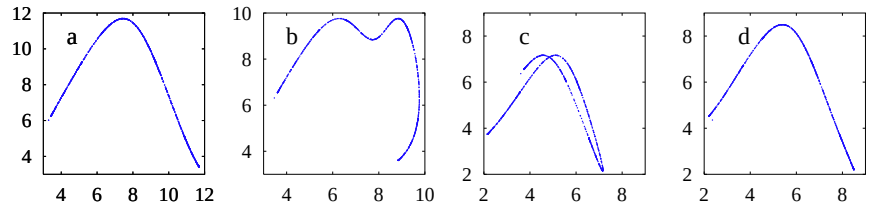



**Figure 3.6:** Return maps for the $R_n \to R_{n+1}$ radial distance Poincaré sections of figure 3.5. (R. Paškauskas)

*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 3.5, illustrates the "stretch & fold" action of the Rössler flow. To orient yourself, compare this with figure 2.5, and note the different $z$-axis scales. Figure 3.5 assembles these sections into a series of snapshots of the flow. A line segment $[A, B]$, traversing the width of the attractor, starts out close to the $x$-$y$ plane, and after the stretching (a) $\to$ (b) followed by the folding (c) $\to$ (d), the folded segment returns close to the $x$-$y$ plane strongly compressed. In one Poincaré return the $[A, B]$ interval is stretched, folded and mapped onto itself, so the flow is* expanding. *It is also* mixing, *as in one Poincaré return the point $C$ from the interior of the attractor is mapped into the outer edge, while the edge point $B$ lands in the interior.*

*Once a particular Poincaré section is picked, we can also exhibit the return map (3.1), as in figure 3.6. Cases (a) and (d) are examples of nice 1-to-1 return maps. However, (b) and (c) appear multimodal and non-invertible, artifacts of projection of a 2-d return map $(R_n, z_n) \to (R_{n+1}, z_{n+1})$ onto a 1-dimensional subspace $R_n \to R_{n+1}$. (Continued in example 4.1)*

fast track:
sect. 3.3, p. 54

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 portrait of the Rössler flow figure 2.5 gives us no sense of the thickness of the attractor, we see clearly in the figure 3.5 Poincaré sections that even though the return map is 2-$d$ $\to$ 2-$d$, the flow contraction is so strong that for all practical purposes it renders the return map 1-dimensional.

### 3.1.1 What's 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, so it would be stupid to keep recomputing them over and over. We would rather replace the whole continuous family of solutions by one.

A smart way to do would be to replace dynamics $(\mathcal{M}, f)$ by dynamics on the *quotient state space* $(\mathcal{M}/G, \tilde{f})$. We will discuss this in chapter 9, but in general constructing explicit quotient state space flow $\tilde{f}$ appears either difficult, or not [chapter 9]

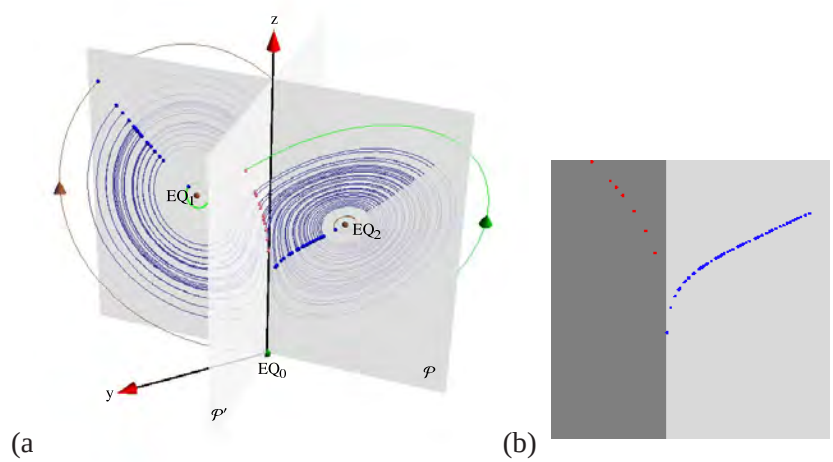



**Figure 3.7:** (a) Lorenz flow figure 2.4 cut by $y = x$ Poincaré section plane $\mathcal{P}$ through the $z$ axis and both $EQ_{1,2}$ equilibria. Points where flow pierces into section are marked by dots. To aid visualization of the flow near the $EQ_0$ equilibrium, the flow is cut by the second Poincaré 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 10.7 (b). (E. Siminos)

appreciated enough to generate much readable literature, or perhaps impossible. So one resorts to method of sections.

Time evolution itself is a 1-parameter abelian Lie group, albeit a highly nontrivial one (otherwise this book would not be much of a doorstop). The invariants of the flow are its infinite-time orbits; particularly useful invariants are compact orbits $\mathcal{M}_p \subset \mathcal{M}$, 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 and its symmetries.

Choice of this one 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, increases the dimensionality of the state space by additional constraints (see sect. 12.4). It 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.

This said, there are a few rules of thumb to follow: (a) You can pick as many sections as convenient. (b) For ease of computation, pick linear sections (3.6) if you can. (c) If equilibria play important role in organizing a flow, pick sections that go through them (see example 3.5). (c) If you have a global discrete or continuous symmetry, pick sections left invariant by the symmetry (see example 9.2). [chapter 9] (d) 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 orthogonal to the flow (e) If you have another rule of thumb dear to you, let us know.

**Example 3.5** ***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 3.7 is a natural choice of a Poincaré section of the Lorenz flow of figure 2.4, as it contains all three equilibria, $x_{EQ0} = (0, 0, 0)$ and the (2.13) pair $EQ_{1,2}$. A section has to be supplemented with the orientation condition (3.4): here points where flow pierces* into *the section are marked by dots.*

*$EQ_{1,2}$ 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 3.7 (b), capture the whole flow near $EQ_0$. In contrast to Rössler sections of figure 3.5, these appear very singular. We explain this singularity in example 4.6, and postpone construction of a Poincaré return map to example 9.2.*

*(E. Siminos and J. Halcrow)*

## 3.2 Constructing a Poincaré section

For almost any flow of physical interest a Poincaré section is not available in analytic form. We describe now a numerical method for determining a Poincaré section.

[remark 3.1]

Consider the system (2.6) of ordinary differential equations in the vector variable $x = (x_1, x_2, \ldots, x_d)$

$$\frac{dx_i}{dt} = v_i(x,t)\,, \tag{3.10}$$

where the flow velocity $v$ is a vector function of the position in state space $x$ and the time $t$. In general, $v$ cannot be integrated analytically, so we will have to 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 (3.6).

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. In order to land exactly on the Poincaré section one often interpolates the intersection point from the two trajectory points on either side of the hypersurface. However, there is a better way.

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

$$\frac{dx_k}{dx_1}\frac{dx_1}{dt} = \frac{dx_k}{dx_1}v_1(x,t) = v_k(x,t) \tag{3.11}$$

we can rewrite the equations of motion (3.10) as

$$\frac{dt}{dx_1} = \frac{1}{v_1}\,, \quad \cdots\,, \quad \frac{dx_d}{dx_1} = \frac{v_d}{v_1}\,. \tag{3.12}$$

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 (3.6). This is the end point of the integration, with no need for any interpolation or backtracking to the surface of 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 3.6** ***Computation of Rössler flow Poincaré sections.*** *Poincaré sections of figure 3.5 are defined by the fixing angle $U(x) = \theta - \theta_0 = 0$. Convert Rössler equation (2.17) to cylindrical coordinates:*

$$\begin{aligned}
\dot{r} &= v_r = -z\cos\theta + ar\sin^2\theta \\
\dot{\theta} &= v_\theta = 1 + \frac{z}{r}\sin\theta + \frac{a}{2}\sin 2\theta \\
\dot{z} &= v_z = b + z(r\cos\theta - c)\,.
\end{aligned} \tag{3.13}$$

*In principle one should use the equilibrium $x_+$ from (2.18) as the origin, and its eigenvectors as the coordinate frame, but here original coordinates suffice, as for parameter values (2.17), and $(x_0, y_0, z_0)$ sufficiently far away from the inner equilibrium, $\theta$ increases monotonically with time. Integrate*

$$\frac{dr}{d\theta} = v_r/v_\theta\,,\ \frac{dt}{d\theta} = 1/v_\theta\,,\ \frac{dz}{d\theta} = v_z/v_\theta \tag{3.14}$$

*from $(r_n, \theta_n, z_n)$ to the next Poincaré section at $\theta_{n+1}$, and switch the integration back to $(x, y, z)$ coordinates.* *(Radford Mitchell, Jr.)*

## 3.3 Maps

Though we have motivated discrete time dynamics by considering sections of a continuous flow, there are many settings in which dynamics is inherently discrete, and naturally described by repeated iterations of the same map

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

or sequences of consecutive applications of a finite set of maps,

$$\{f_A, f_B, \ldots f_Z\} : \mathcal{M} \to \mathcal{M}\,, \tag{3.15}$$

for example maps relating different sections among a set of Poincaré sections. The discrete 'time' is then an integer, the number of applications of a map. As writing out formulas involving repeated applications of a set of maps explicitly can be awkward, we streamline the notation by denoting a map composition by '$\circ$'

$$f_Z(\cdots f_B(f_A(x)))\cdots) = f_Z \circ \cdots f_B \circ f_A(x)\,, \tag{3.16}$$

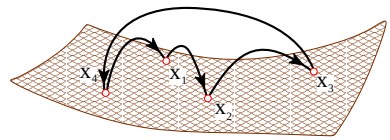



**Figure 3.8:** A flow $x(t)$ of figure 3.1 represented by a Poincaré return map that maps points in the Poincaré section $\mathcal{P}$ as $x_{n+1} = f(x_n)$. In this example the orbit of $x_1$ consists of the four cycle points $(x_1, x_2, x_3, x_4)$

and the $n$th iterate of map $f$ by

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

[section 2.1]

The *trajectory* of $x$ is now the discrete set of points

$$\left\{x, f(x), f^2(x), \ldots, f^n(x)\right\},$$

and the *orbit* of $x$ is the subset of all points of $\mathcal{M}$ that can be reached by iterations of $f$. For example, the orbit of $x_1$ in figure 3.8 is the 4-cycle $(x_1, x_2, x_3, x_4)$.

The functional form of such Poincaré return maps $P$ as figure 3.6 can be approximated by tabulating the results of integration of the flow from $x$ to the first Poincaré section return for many $x \in \mathcal{P}$, and constructing a function that interpolates through these points. If we find a good approximation to $P(x)$, we can get rid of numerical integration altogether, by replacing the continuous time trajectory $f^t(x)$ by iteration of the Poincaré return map $P(x)$. Constructing accurate $P(x)$ for a given flow can be tricky, but we can already learn much from approximate Poincaré return maps. Multinomial approximations

$$P_k(x) = a_k + \sum_{j=1}^{d} b_{kj} x_j + \sum_{i,j=1}^{d} c_{kij} x_i x_j + \ldots, \qquad x \in \mathcal{P} \tag{3.17}$$

to Poincaré return maps

$$\begin{pmatrix} x_{1,n+1} \\ x_{2,n+1} \\ \ldots \\ x_{d,n+1} \end{pmatrix} = \begin{pmatrix} P_1(x_n) \\ P_2(x_n) \\ \ldots \\ P_d(x_n) \end{pmatrix}, \qquad x_n, x_{n+1} \in \mathcal{P}$$

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

**Example 3.7** ***Hénon map:*** *The map*

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

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

$$x_{n+1} = 1 - a x_n^2 + b x_{n-1}. \tag{3.19}$$

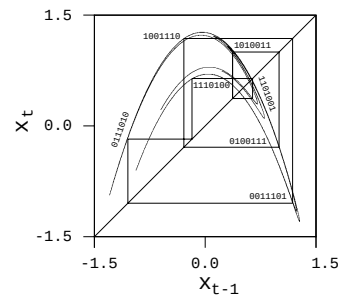



**Figure 3.9:** The strange attractor and an unstable period 7 cycle of the Hénon map (3.18) with $a = 1.4$, $b = 0.3$. The periodic points in the cycle are connected to guide the eye. (K.T. Hansen [8])

*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 3.5. It can be obtained by a truncation of a polynomial approximation (3.17) to a Poincaré return map (3.17) to second order.*

*A quick sketch of the long-time dynamics of such a mapping (an example is depicted in figure 3.9), is obtained by picking an arbitrary starting point and iterating (3.18) on a computer. We plot here the dynamics in the $(x_n, x_{n+1})$ plane, rather than in the $(x_n, y_n)$ plane, because we think of the Hénon map as a model return map $x_n \to x_{n+1}$. 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. 27.1 , and the cycle point labels* 0111010, 1110100, $\cdots$ *in sect. 11.3.* [exercise 3.5]

**Example 3.8** ***Lozi map:*** *Another example frequently employed is the* Lozi map, *a linear, 'tent map' version of the Hénon map (3.18) given by*

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

*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.*

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 3.6. 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 3.5]

**Example 3.9** ***Parabola:*** *For sufficiently large value of the stretching paramater $a$, one iteration of the Hénon map (3.18) stretches and folds a region of the $(x, y)$ plane centered around the origin. The parameter $a$ controls the amount of stretching, while*

*the parameter $b$ controls the thickness of the folded image through the '1-step memory' term $bx_{n-1}$ in (3.19). In figure 3.9 the parameter $b$ is rather large, $b = 0.3$, so the attractor is rather thick, with the transverse fractal structure clearly visible. For vanishingly small $b$ the Hénon map reduces to the 1-dimensional quadratic map*

$$x_{n+1} = 1 - ax_n^2 \,. \tag{3.21}$$

[exercise 3.7]

*By setting $b = 0$ we lose determinism, as on reals the inverse of map (3.21) has two 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.*

As we shall see in sect. 10.2.1, 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 H.8]

## Résumé

In recurrent dynamics a trajectory exits a region in state space and then reenters it infinitely often, with a finite mean return time. If the orbit is periodic, it returns after a full period. So, on average, nothing much really happens along the trajectory–what is important is behavior of neighboring trajectories transverse to the flow. This observation motivates a replacement of the continuous time flow by iterative mapping, the Poincaré return map.

The visualization of strange attractors is greatly facilitated by a felicitous choice of Poincaré sections, and the reduction of flows to Poincaré 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 8. As we shall show in chapter 6, further simplification of a Poincaré return map, or any nonlinear map, can be attained through rectifying these maps locally by means of smooth conjugacies.

[chapter 8]
[chapter 6]

## Commentary

**Remark 3.1** Determining a Poincaré section. 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. The trick described in sect. 3.2 is due to Hénon [5, 6, 7].

**Remark 3.2** 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 [1], see ref. [1] and remark 2.2.

Y. Pomeau's studies of the Lorenz attractor on an analog computer, and his insights into its stretching and folding dynamics motivated Hénon [2] to introduce the Hénon map in 1976. Hénon's and Lorenz's original papers can be found in reprint collections refs. [3, 4]. They are a pleasure to read, and are still the best introduction to the physics motivating such models. A detailed description of the dynamics of the Hénon map is given by Mira and coworkers [8], as well as very many other authors.

The Lozi map [10] is particularly convenient in investigating the symbolic dynamics of 2-*d* mappings. Both the Lorenz and Lozi systems are uniformly smooth systems with singularities. The continuity of measure for the Lozi map was proven by M. Misiurewicz [11], and the existence of the SRB measure was established by L.-S. Young. [section 14.1]

**Remark 3.3** Grasshoppers *vs.* butterflies. The 'sensitivity to initial conditions' was discussed by Maxwell, 30 years later by Poincaré. In weather prediction, the Lorentz' 'Butterfly Effect' started its journey in 1898, as a 'Grasshopper Effect' in a book review by W. S. Franklin [1]. In 1963 Lorenz ascribed a 'seagull effect' to an unnamed meteorologist, and in 1972 he repackaged it as the 'Butterfly Effect'.

## 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. As the Rössler flow state space is $3D$, the flow maps onto a $2D$ Poincaré section. Do you see that in your numerical results? How good an approximation would a replacement of the return map for this section by a 1-dimensional map be? More precisely, estimate the thickness of the strange attractor. (continued as exercise 4.4)

(R. Paškauskas)

3.2. **A return Poincaré map for the Rössler flow.** (continuation of exercise 3.1) That Poincaré return maps of figure 3.6 appear multimodal and non-invertible is an artifact of projections of a 2-dimensional return map $(R_n, z_n) \to (R_{n+1}, z_{n+1})$ onto a 1-dimensional subspace $R_n \to R_{n+1}$.

Construct a genuine $s_{n+1} = f(s_n)$ return map by parametrazing points on a Poincaré section of the attractor figure 3.5 by a Euclidean length $s$ computed curvilinearly along the attractor section.

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 12.1 (b). (P. Cvitanović)

3.3. **Arbitrary Poincaré sections.** We will generalize the construction of Poincaré sections so that they can have any shape, 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 equations as

$$\frac{dx_k}{ds} = \kappa f_k \,, \tag{3.22}$$

with $dt/ds = \kappa$, and choosing $\kappa$ 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

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

How can this be used to find a Poincaré section?

3.4. **Classical collinear helium dynamics.**

(continuation of exercise 2.10) Make a Poincaré surface of section by plotting $(r_1, p_1)$ whenever $r_2 = 0$: Note that for $r_2 = 0$, $p_2$ is already determined by (7.6). Compare your results with figure 6.3 (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

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

3.6. **How strange is the Hénon attractor?**

(a) Iterate numerically 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 $a = 1.4$, $b = 0.3$ . Would you describe the result as a 'strange attractor'? Why?

(b) Now check how robust the Hénon attractor is by iterating a slightly different Hénon map, with $a = 1.39945219$, $b = 0.3$. Keep at it until the 'strange' attractor vanishes like the smile of the Chesire cat. What replaces it? Would you describe the result as a 'strange attractor'? Do you still have confidence in your own claim for the part (a) of this exercise?

3.7. **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?

## References

[3.1] W. S. Franklin, "New Books," *Phys. Rev.* **6**, 173 (1898); see www.ceafinney.com/chaos.

[3.2] M. Hénon, *Comm. Math. Phys.* **50**, 69 (1976).

[3.3] *Universality in Chaos*, P. Cvitanović, ed., (Adam Hilger, Bristol 1989).

[3.4] Bai-Lin Hao, *Chaos* (World Scientific, Singapore, 1984).

[3.5] M. Hénon, "On the numerical computation of Poincaré maps," *Physica* **D 5**, 412 (1982).

[3.6] N.B. Tufillaro, T.A. Abbott, and J.P. Reilly, *Experimental Approach to Nonlinear Dynamics and Chaos* (Addison Wesley, Reading MA, 1992).

[3.7] Bai-Lin Hao, *Elementary symbolic dynamics and chaos in dissipative systems* (World Scientific, Singapore, 1989).

[3.8] C. Mira, *Chaotic Dynamics–From one dimensional endomorphism to two dimensional diffeomorphism*, (World Scientific, Singapore, 1987).

[3.9] I. Gumowski and C. Mira, *Recurrances and Discrete Dynamical Systems* (Springer-Verlag, Berlin 1980).

[3.10] R. Lozi, *J. Phys. (Paris) Colloq.* **39**, 9 (1978).

[3.11] M. Misiurewicz, *Publ. Math. IHES* **53**, 17 (1981).

[3.12] D. Fournier, H. Kawakami and C. Mira, *C.R. Acad. Sci. Ser. I*, **298**, 253 (1984); **301**, 223 (1985); **301**, 325 (1985).

[3.13] M. Benedicks and L.-S. Young,

*Ergodic Theory & Dynamical Systems* **12**, 13–37 (1992).

# Chapter 4

# Local stability

(R. Mainieri and P. Cvitanović)

So far we have concentrated on description of 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 describe 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 16, in hyperbolic systems what matters are the expanding directions. The repercussion 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.

## 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 understood by considering a trajectory originating near $x_0 = x(0)$ with an initial infinitesimal displacement $\delta x(0)$, and letting the flow transport the displacement $\delta x(t)$ along the trajectory $x(x_0, t) = f^t(x_0)$.

### 4.1.1 Instantaneous shear

The system of linear *equations of variations* for the displacement of the infinitesimally close neighbor $x + \delta x$ follows from the flow equations (2.6) by Taylor

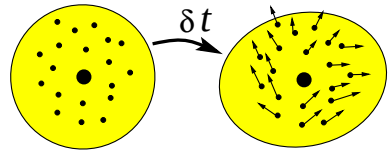



**Figure 4.1:** A swarm of neighboring points of $x(t)$ is instantaneously sheared by the action of the stability matrix $A$ - a bit hard to draw.

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 displacement $\delta x$ is thus transported along the trajectory $x(x_0, t)$, with time variation given by

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

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 $x(x_0, t) \to x(t) \to x$, $\delta x_i(x_0, t) \to \delta x_i(t) \to \delta x$ in what follows. Taken together, the set of equations

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

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

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

describes the instantaneous rate of shearing of the infinitesimal neighborhood of $x(t)$ by the flow, figure 4.1.

**Example 4.1** ***Rössler and Lorenz flows, linearized:*** *For the Rössler (2.17) and Lorenz (2.12) flows the stability matrices are, respectively*

$$A_{Ross} = \begin{pmatrix} 0 & -1 & -1 \\ 1 & a & 0 \\ z & 0 & x-c \end{pmatrix}, \qquad A_{Lor} = \begin{pmatrix} -\sigma & \sigma & 0 \\ \rho - z & -1 & x \\ y & x & -b \end{pmatrix}. \tag{4.4}$$

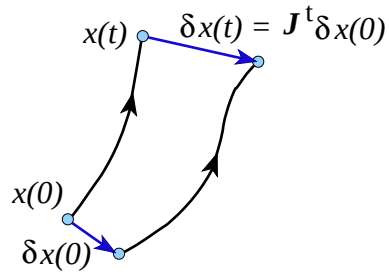



**Figure 4.2:** The fundamental matrix $J^t$ maps an infinitesimal displacement at $x_0$ into a displacement rotated and sheared by the linearized flow fundamental matrix $J^t(x_0)$ finite time $t$ later.

### 4.1.2 Linearized flow

Major combat operations in Iraq have ended.
— President G. W. Bush, May 1, 2003

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

$$f_i^t(x_0 + \delta x) = f_i^t(x_0) + \sum_j \frac{\partial f_i^t(x_0)}{\partial x_{0\,j}} \delta x_j + \cdots , \tag{4.5}$$

one finds that the linearized neighborhood is transported by

$$\delta x(t) = J^t(x_0)\delta x_0 , \qquad J_{ij}^t(x_0) = \left.\frac{\partial x_i(t)}{\partial x_j}\right|_{x=x_0} . \tag{4.6}$$

This Jacobian matrix has inherited the name *fundamental solution matrix* or simply *fundamental matrix* from the theory of linear ODEs. 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. $J$ describes the deformation of an infinitesimal neighborhood at finite time $t$ in the co-moving frame of $x(t)$.

As this is a deformation in the linear approximation, one can think of it as a linear deformation of an infinitesimal sphere enveloping $x_0$ into an ellipsoid around $x(t)$, described by the eigenvectors and eigenvalues of the fundamental matrix of the linearized flow, figure 4.2. Nearby trajectories separate along the *unstable directions*, approach each other along the *stable directions*, and change their distance along the *marginal directions* at a rate slower than exponential, corresponding to the eigenvalues of the fundamental matrix with magnitude larger than, smaller than, or equal 1. In the literature adjectives *neutral* or *indifferent* are often used instead of 'marginal,' (attracting) stable directions are sometimes called 'asymptotically stable,' and so on.

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

$$f^{\delta t+t}(x_0) = \int_0^{\delta t+t} d\tau\, v(x(\tau)) = \delta t\, v(x(t)) + f^t(x_0) .$$

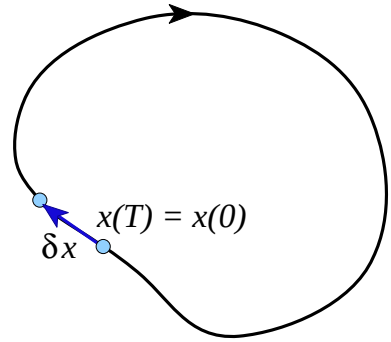



**Figure 4.3:** For a periodic orbit $p$, any two points along the cycle are mapped into themselves after one cycle period $T$, hence $\delta x = v(x_0)\delta t$ is mapped into itself by the cycle fundamental matrix $J_p$.

Expanding both sides of $f^t(f^{\delta t}(x_0)) = f^{\delta t}(f^t(x_0))$, keeping the leading term in $\delta t$, and using the definition of the fundamental matrix (4.6), we observe that $J^t(x_0)$ transports the velocity vector at $x_0$ to the velocity vector at $x(t)$ at time $t$:

$$v(x(t)) = J^t(x_0)\, v(x_0)\,. \tag{4.7}$$

In nomenclature of page 63, the fundamental matrix maps the initial, Lagrangian coordinate frame into the current, Eulerian coordinate frame.

The velocity at point $x(t)$ in general 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 fundamental matrix computed for an arbitrary segment of an arbitrary trajectory has no invariant meaning.

As the eigenvalues of finite time $J^t$ have invariant meaning only for periodic orbits, we postpone their interpretation to chapter 5. However, already at this stage we see that if the orbit is periodic, $x(T_p) = x(0)$, at any point along cycle $p$ the velocity $v$ is an eigenvector of the fundamental matrix $J_p = J^{T_p}$ with a unit eigenvalue,

$$J_p(x)\, v(x) = v(x)\,, \qquad x \in p\,. \tag{4.8}$$

Two successive points along the cycle separated by $\delta x_0$ have the same separation after a completed period $\delta x(T_p) = \delta x_0$, see figure 4.3, hence eigenvalue 1.

As we started by assuming that we know the equations of motion, 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 (4.6).

Our next task is to relate the stability matrix $A$ to fundamental matrix $J^t$. On the level of differential equations the relation follows by taking the time derivative of (4.6) and replacing $\dot{\delta x}$ by (4.2)

$$\dot{\delta x}(t) = \dot{J}^t\, \delta x_0 = A\delta x(t) = AJ^t\, \delta x_0\,.$$

Hence the $d^2$ matrix elements of fundamental matrix satisfy the linearized equation (4.1)

$$\frac{d}{dt} J^t(x) = A(x)\, J^t(x)\,, \quad \text{initial condition} \;\; J^0(x) = \mathbf{1}\,. \tag{4.9}$$

Given a numerical routine for integrating the equations of motion, evaluation of the fundamental matrix requires minimal additional programming effort; one simply extends the $d$-dimensional integration routine and integrates concurrently with $f^t(x)$ the $d^2$ elements of $J^t(x)$.

The qualifier 'simply' is perhaps too glib. Integration will work for short finite times, but for exponentially unstable flows one quickly runs into numerical over- and/or underflow problems, so further thought will have to go into implementation this calculation.

So now we know how to compute fundamental 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 7, p. 108

And yet... there are mopping up operations left to do. We persist until we derive the integral formula (4.43) for the fundamental matrix, an analogue of the finite-time "Green 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 *non*linear dynamics?), offers a firm stepping stone on the way to understanding nonlinear flows. If you know your eigenvalues and eigenvectors, you may prefer to fast forward here.

fast track:
sect. 4.3, p. 71

## 4.2 Linear flows

Diagonalizing the matrix: that's the key to the whole thing.
— Governor Arnold Schwarzenegger

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 (2.6) are written in terms of a vector $x$ and a constant stability matrix $A$ as

$$\dot{x} = v(x) = Ax\,. \tag{4.10}$$

Solving this equation means finding the state space trajectory

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

passing through the 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, which means that one can find a basis of $d$ linearly independent solutions.

How do we solve the linear differential equation (4.10)? If instead of a matrix equation we have a scalar one, $\dot{x} = \lambda x$, the solution is

$$x(t) = e^{t\lambda} x_0 \,. \tag{4.11}$$

In order to solve the $d$-dimensional matrix case, it is helpful to rederive the solution (4.11) by studying what happens for a short time step $\delta t$. If at time $t = 0$ the position is $x(0)$, then

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

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

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

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 $m \to \infty$, the term in parentheses converges to $e^{t\lambda}$. Consider now the matrix version of equation (4.12):

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

A representative point $x$ is now a vector in $\mathbb{R}^d$ acted on by the matrix $A$, as in (4.10). Denoting by $\mathbf{1}$ the identity matrix, and repeating the steps (4.12) and (4.13) we obtain Euler's formula for the exponential of a matrix:

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

We will find this definition the exponential of a matrix helpful in the general case, where the matrix $A = A(x(t))$ varies along a trajectory.

How do we compute the exponential (4.15)?

fast track:
sect. 4.3, p. 71

**Example 4.2** ***Fundamental matrix eigenvalues, diagonalizable case:*** *Should we be so lucky that $A = A_D$ happens to be a diagonal matrix with eigenvalues $(\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(d)})$, the exponential is simply*

$$J^t = e^{tA_D} = \begin{pmatrix} e^{t\lambda^{(1)}} & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & e^{t\lambda^{(d)}} \end{pmatrix}. \tag{4.16}$$

*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)):*

[exercise 4.2]

$$J^t = e^{tA} = Ue^{tA_D}U^{-1}. \tag{4.17}$$

*The action of both $A$ and $J$ is very simple; the axes of orthogonal coordinate system where $A$ is diagonal are also the eigen-directions of both $A$ and $J^t$, and under the flow the neighborhood is deformed by a multiplication by an eigenvalue factor for each coordinate axis.*

In general $J^t$ is neither diagonal, nor diagonalizable, nor constant along the trajectory. As any matrix, $J^t$ can also be expressed in the singular value decomposition form

$$J = UDV^T$$

where $D$ is diagonal, and $U$, $V$ are orthogonal matrices. The diagonal elements $\sigma_1, \sigma_2, \ldots, \sigma_d$ of $D$ are called the *singular values* of $J$, namely the square root of the eigenvalues of $J^\dagger J$, which is a Hermitian, positive semi-definite matrix (and thus admits only real, non-negative eigenvalues). From a geometric point of view, when all singular values are non-zero, $J$ maps the unit sphere into an ellipsoid: the singular values are then the lengths of the semiaxes of this ellipsoid.

[section 5.1.2]

We recapitulate the basic facts of linear algebra in appendix B. A 2-*d* example serves well to highlight the most important types of linear flows:

**Example 4.3** ***Linear stability of 2-d flows:*** *For a 2-d 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 a 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 $\lambda^{(1)} > 0$, $\lambda^{(2)} < 0$, $x_1$ grows exponentially with time, and $x_2$ contracts exponentially. This behavior, called a* saddle*, is sketched in figure 4.4, 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 contracts into $(0, 0)$ when the $\mu < 0$, whereas the phase velocity $\omega$ controls its oscillations.*

*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. (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.*

*We sketch the full set of possibilities in figures 4.4 and 4.5, and work out in detail the most important cases in appendix B, example B.2.*

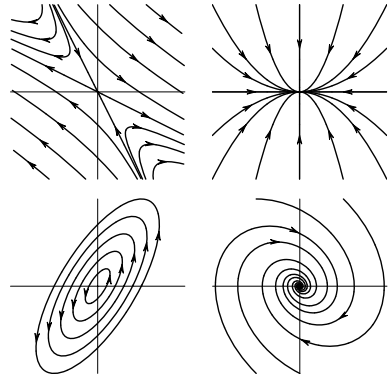


**Figure 4.4:** Streamlines for several typical 2-dimensional flows: saddle (hyperbolic), in node (attracting), center (elliptic), in spiral.

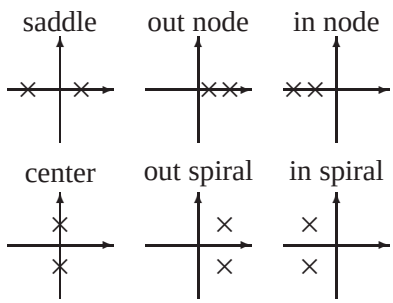



**Figure 4.5:** Qualitatively distinct types of exponents of a [2×2] fundamental matrix.

### 4.2.1 Eigenvalues, multipliers - a notational interlude

Throughout this text the symbol $\Lambda_k$ will always denote the $k$th *eigenvalue* (in literature sometimes referred to as the *multiplier* or *Floquet!multiplier*) of the finite time fundamental matrix $J^t$. Symbol $\lambda^{(k)}$ will be reserved for the $k$th *Floquet* or *characteristic* exponent, or *characteristic value*, with real part $\mu^{(k)}$ and phase $\omega^{(k)}$:

$$\Lambda_k = e^{t\lambda^{(k)}} = e^{t(\mu^{(k)}+i\omega^{(k)})} . \tag{4.18}$$

$J^t(x_0)$ depends on the initial point $x_0$ and the elapsed time $t$. For notational brevity we tend to omit this dependence, but in general

$$\Lambda = \Lambda_k = \Lambda_k(x_0,t)\,,\ \lambda = \lambda^{(k)}(x_0,t)\,,\ \omega = \omega^{(k)}(x_0,t)\,,\cdots \text{ etc.}\,,$$

depend on both the trajectory traversed and the choice of coordinates.

However, as we shall see in sect. 5.2, if the stability matrix $A$ or the fundamental matrix $J$ is computed on a flow-invariant set $\mathcal{M}_p$, such as an equilibrium $q$ or a periodic orbit $p$ of period $T_p$,

$$A_q = A(x_q)\,, \qquad J_p(x) = J^{T_p}(x)\,, \quad x \in \mathcal{M}_p\,, \tag{4.19}$$

($x$ is any point on the cycle) its eigenvalues

$$\lambda_q^{(k)} = \lambda^{(k)}(x_q)\,,\ \Lambda_{p,k} = \Lambda_k(x,T_p)$$

are flow-invariant, independent of the choice of coordinates and the initial point in the cycle $p$, so we label them by their $q$ or $p$ label.

We number eigenvalues $\Lambda_k$ in order of decreasing magnitude

$$|\Lambda_1| \geq |\Lambda_2| \geq \ldots \geq |\Lambda_d| \,. \tag{4.20}$$

Since $|\Lambda_j| = e^{t\mu^{(j)}}$, this is the same as labeling by

$$\mu^{(1)} \geq \mu^{(2)} \geq \ldots \geq \mu^{(d)} \,. \tag{4.21}$$

In dynamics the expanding directions, $|\Lambda_e| > 1$, have to be taken care of first, while the contracting directions $|\Lambda_c| < 1$ tend to take care of themselves, hence the ordering by decreasing magnitude is the natural one.

### 4.2.2 Yes, but how do you really do it?

Economical description of neighborhoods of equilibria and periodic orbits is afforded by projection operators

$$\mathbf{P}_i = \prod_{j\neq i} \frac{\mathbf{M} - \lambda^{(j)}\mathbf{1}}{\lambda^{(i)} - \lambda^{(j)}} \,, \tag{4.22}$$

where matrix $\mathbf{M}$ is typically either equilibrium stability matrix $A$, or periodic orbit fundamental matrix $\hat{J}$ restricted to a Poincaré section, as in (4.55). While usually not phrased in language of projection operators, the requisite linear algebra is standard, and relegated here to appendix B.

Once the distinct non-zero eigenvalues $\{\lambda^{(i)}\}$ are computed, projection operators are polynomials in $\mathbf{M}$ which need no further diagonalizations or orthogonalizations. For each distinct eigenvalue $\lambda^{(i)}$ of $\mathbf{M}$, the colums/rows of $\mathbf{P}_i$

$$(\mathbf{M} - \lambda^{(j)}\mathbf{1})\mathbf{P}_j = \mathbf{P}_j(\mathbf{M} - \lambda^{(j)}\mathbf{1}) = 0 \,, \tag{4.23}$$

are the right/left eigenvectors $\mathbf{e}^{(k)}$, $\mathbf{e}_{(k)}$ of $\mathbf{M}$ which (provided $\mathbf{M}$ is not of Jordan type) span the corresponding linearized subspace, and are a convenient starting seed for tracing out the global unstable/stable manifolds.

Matrices $\mathbf{P}_i$ are *orthogonal* and *complete*:

$$\mathbf{P}_i\mathbf{P}_j = \delta_{ij}\mathbf{P}_j \,, \quad \text{(no sum on } j\text{)} \,, \qquad \sum_{i=1}^{r} \mathbf{P}_i = \mathbf{1} \,. \tag{4.24}$$

with the dimension of the $i$th subspace given by $d_i = \mathrm{tr}\,\mathbf{P}_i$ . Completeness relation substituted into $\mathbf{M} = \mathbf{M}\,\mathbf{1}$ yields

$$\mathbf{M} = \lambda^{(1)}\mathbf{P}_1 + \lambda^{(2)}\mathbf{P}_2 + \cdots + \lambda^{(r)}\mathbf{P}_r\,. \tag{4.25}$$

As any matrix function $f(\mathbf{M})$ takes the scalar value $f(\lambda^{(i)})$ on the $\mathbf{P}_i$ subspace, $f(\mathbf{M})\mathbf{P}_i = f(\lambda^{(i)})\mathbf{P}_i$ , it is easily evaluated through its *spectral decomposition*

$$f(\mathbf{M}) = \sum_i f(\lambda^{(i)})\mathbf{P}_i\,. \tag{4.26}$$

As $\mathbf{M}$ 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). That is not surprising, but also 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, $\{\mathbf{e}^{(k)}, \mathbf{e}^{(k+1)}\} \to \{\mathrm{Re}\,\mathbf{e}^{(k)}, \mathrm{Im}\,\mathbf{e}^{(k)}\}$, or, in terms of projection operators:

$$\mathbf{P}_k = \frac{1}{2}(\mathbf{R} + i\mathbf{Q})\,, \qquad \mathbf{P}_{k+1} = \mathbf{P}_k^*\,,$$

where $\mathbf{R} = \mathbf{P}_k + \mathbf{P}_{k+1}$ is the subspace decomposed by the $k$th complex eigenvalue pair, and $\mathbf{Q} = (\mathbf{P}_k - \mathbf{P}_{k+1})/i$, both matrices with real elements. Substitution

$$\begin{pmatrix} \mathbf{P}_k \\ \mathbf{P}_{k+1} \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}\begin{pmatrix} \mathbf{R} \\ \mathbf{Q} \end{pmatrix}\,,$$

brings the $\lambda^{(k)}\mathbf{P}_k + \lambda^{(k+1)}\mathbf{P}_{k+1}$ complex eigenvalue pair in the spectral decomposition (4.25) into the real form,

$$\begin{pmatrix} \mathbf{P}_k & \mathbf{P}_{k+1} \end{pmatrix}\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^* \end{pmatrix}\begin{pmatrix} \mathbf{P}_k \\ \mathbf{P}_{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{R} & \mathbf{Q} \end{pmatrix}\begin{pmatrix} \mu & -\omega \\ \omega & \mu \end{pmatrix}\begin{pmatrix} \mathbf{R} \\ \mathbf{Q} \end{pmatrix}\,, \tag{4.27}$$

where we have dropped the superscript $^{(k)}$ for notational brevity.

To summarize, spectrally decomposed matrix $\mathbf{M}$ (4.25) 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.

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.

## 4.3 Stability of flows

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

$$x(t) = f^t(x_0)\,, \quad \delta x(x_0,t) \;=\; J^t(x_0)\delta x(x_0,0)\,. \tag{4.28}$$

The equations are linear, so we should be able to integrate them–but in order to make sense of the answer, we derive it step by step.

### 4.3.1 Stability of equilibria

For a start, consider the case where $x$ is an equilibrium point (2.8). 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,

$$f^t(x) = x_q + e^{At}(x - x_q) + \cdots\,, \tag{4.29}$$

we verify that the simple formula (4.15) applies also to the fundamental matrix of an equilibrium point,

$$J^t(x_q) = e^{A_q t}\,, \qquad A_q = A(x_q)\,. \tag{4.30}$$

**Example 4.4** ***In-out spirals.*** *Consider a 2-d equilibrium whose stability eigenvalues* $\{\lambda^{(1)},\lambda^{(2)}\} = \{\mu + i\omega, \mu - i\omega\}$ *form a complex conjugate pair. The corresponding complex eigenvectors can be replaced by their real and imaginary parts,* $\{\mathbf{e}^{(1)},\mathbf{e}^{(2)}\} \to \{\mathrm{Re}\,\mathbf{e}^{(k)}, \mathrm{Im}\,\mathbf{e}^{(k)}\}$. *The 2-d real representation (4.27),*

$$\begin{pmatrix} \mu & -\omega \\ \omega & \mu \end{pmatrix} = \mu \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \omega \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

*consists of the identity and the generator of* $SO(2)$ *rotations. Trajectories* $\mathbf{x}(t) = J^t\,\mathbf{x}(0)$, *where*

$$J^t = e^{A_q t} = e^{t\mu} \begin{pmatrix} \cos\,\omega t & -\sin\,\omega t \\ \sin\,\omega t & \cos\,\omega t \end{pmatrix}\,, \tag{4.31}$$

*spiral in/out around* $(x,y) = (0,0)$, *see figure 4.4, with the rotation period* $T$, *and contraction/expansion radially by the multiplier* $\Lambda_{radial}$, *and by the multiplier* $\Lambda_j$ *along the* $\mathbf{e}^{(j)}$ *eigendirection per a turn of the spiral:*

[exercise B.1]

$$T = 2\pi/\omega\,, \qquad \Lambda_{radial} = e^{T\mu}\,, \qquad \Lambda_j = e^{T\mu^{(j)}}\,. \tag{4.32}$$

*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* $10^{-2}T$*).* $\Lambda_j$ *multipliers give us estimates of strange-set thickness.*

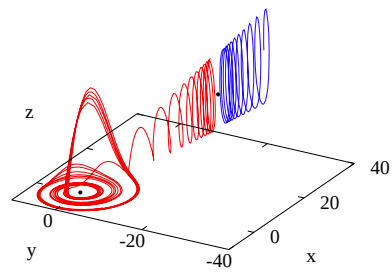



**Figure 4.6:** Two trajectories of the Rössler flow initiated in the neighborhood of the '+' or 'outer' equilibrium point (2.18). (R. Paškauskas)

**Example 4.5** ***Stability of equilibria of the Rössler flow.*** *The Rösler system (2.17) has two equilibrium points (2.18), 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 now yield quite detailed information about the flow. Figure 4.6 shows two trajectories which start in the neighborhood of the '+' equilibrium point. Trajectories to the right of the outer 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 outer equilibrium point thus serves as a attraction basin boundary. Consider now the eigenvalues of the two equilibria* [exercise 4.4] [exercise 2.8]

$$\begin{aligned}(\mu_-^{(1)}, \mu_-^{(2)} \pm i\,\omega_-^{(2)}) &= \quad (-5.686, \qquad 0.0970 \pm i\,0.9951\,) \\ (\mu_+^{(1)}, \mu_+^{(2)} \pm i\,\omega_+^{(2)}) &= \quad (\,0.1929, \quad -4.596\times 10^{-6} \pm i\,5.428\,)\end{aligned} \tag{4.33}$$

**Outer equilibrium:** *The* $\mu_+^{(2)} \pm i\,\omega_+^{(2)}$ *complex eigenvalue pair implies that that neighborhood of the outer equilibrium point rotates with angular period* $T_+ \approx \left|2\pi/\omega_+^{(2)}\right| = 1.1575$*. The multiplier by which a trajectory that starts near the '+' equilibrium point contracts in the stable manifold plane is the excrutiatingly slow* $\Lambda_2^+ \approx \exp(\mu_+^{(2)} T_+) = 0.9999947$ *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 neighborhood of the '−' equilibrium point rotates with angular period* $T_- \approx \left|2\pi/\omega_-^{(2)}\right| = 6.313$*, slightly faster than the harmonic oscillator estimate in (2.14). The multiplier by which a trajectory that starts near the '−' equilibrium point spirals away per one rotation is* $\Lambda_{radial} \approx \exp(\mu_-^{(2)} T_-) = 1.84$*. The* $\mu_-^{(1)}$ *eigenvalue is essentially the* $z$ *expansion correcting parameter* $c$ *introduced in (2.16). 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* $\Lambda_1$ *eigen-direction. After one Poincaré return the interval is of 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 Poincaré return map is invertible.* *(R. Paškauskas)*

**Example 4.6** ***Stability of Lorenz flow equilibria:*** *(Continued from example 3.5.) A glance at figure 3.7 suggests that the flow is organized by its 3 equilibria, so lets have a closer look at their stable/unstable manifolds.*

*Lorenz flow is volume contracting (4.47),*

$$\partial_i v_i = \sum_{i=1}^{3} \lambda^{(i)}(x,t) = -\sigma - b - 1\,, \tag{4.34}$$

*at a constant, coordinate- and* $\rho$*-independent rate, set by Lorenz to* $\partial_i v_i = -13.66$ *.*

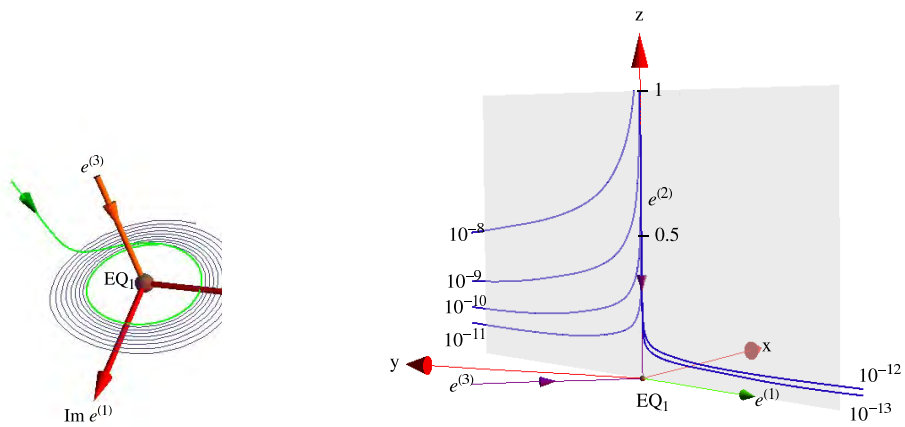



**Figure 4.7:** (a) A perspective view of the linearized Lorenz flow near $EQ_1$ equilibrium, see figure 3.7 (a). The unstable eigenplane of $EQ_1$ is spanned by $\mathrm{Re}\,\mathbf{e}^{(1)}$ and $\mathrm{Im}\,\mathbf{e}^{(1)}$. The stable eigenvector $\mathbf{e}^{(3)}$. (b) Lorenz flow near the $EQ_0$ equilibrium: unstable eigenvector $\mathbf{e}^{(1)}$, stable eigenvectors $\mathbf{e}^{(2)}$, $\mathbf{e}^{(3)}$. Trajectories initiated at distances $10^{-8}\cdots 10^{-12}$, $10^{-13}$ 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 $EQ_1 \to EQ_0$ heteroclinic connection never observed in simulations such as figure 2.4. (E. Siminos; continued in figure 10.7.)

*The symmetry of Lorenz flow leads to a block-diagonal form for the $EQ_0$ equilibrium stability matrix, explicit in (4.4) evaluated at $x_{EQ0} = (0,0,0)$. The $z$-axis is an eigenvector with a contracting eigenvalue $\lambda^{(2)} = -b$. From (4.34) it follows that all $[x,y]$ areas shrink at rate $-\sigma - b$. Indeed, the $[x,y]$ submatrix*

$$A^- = \begin{pmatrix} -\sigma & \sigma \\ \rho & -1 \end{pmatrix} \tag{4.35}$$

*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*

$$\mathbf{P}_i = \frac{A^- - \lambda^{(j)}\mathbf{1}}{\lambda^{(i)} - \lambda^{(j)}} = \frac{1}{\lambda^{(i)} - \lambda^{(j)}} \begin{pmatrix} -\sigma - \lambda^{(j)} & \sigma \\ \rho & -1 - \lambda^{(j)} \end{pmatrix}, \qquad i \neq j \in \{1,3\}\,. \tag{4.36}$$

*$EQ_{1,2}$ equilibria have no symmetry, so their eigenvalues are given by the roots of a cubic equation, the secular determinant $\det(A - \lambda\mathbf{1}) = 0$:*

$$\lambda^3 + \lambda^2(\sigma + b + 1) + \lambda b(\sigma + \rho) + 2\sigma b(\rho - 1) = 0\,. \tag{4.37}$$

*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 2.4.*

*As all numerical plots of the Lorenz flow are here carried out for the Lorenz parameter choice $\sigma = 10, b = 8/3, \rho = 28$, we note the values of these eigenvalues for future reference,*

$$\begin{aligned} EQ_0 :\ (\lambda^{(1)}, \lambda^{(2)}, \lambda^{(3)}) &= (\,11.83\,,\ -2.666,\ \ -22.83\,) \\ EQ_1 :\ (\mu^{(1)} \pm i\,\omega^{(1)}, \lambda^{(3)}) &= (\,0.094 \pm i\,10.19,\ \ -13.85\,)\,, \end{aligned} \tag{4.38}$$

*as well as the rotation period $T_{EQ1} = 2\pi/\omega^{(1)}$ about $EQ_1$, and the associated expansion/contraction multipliers $\Lambda^{(i)} = \exp(\mu^{(j)}T_{EQ1})$ per a spiral-out turn:*

$$T_{EQ1} = 0.6163\,, \qquad (\Lambda^{(1)}, \Lambda^{(3)}) = (\,1.060\,, 1.957\times 10^{-4}\,)\,. \tag{4.39}$$

*We learn that the typical turnover time scale in this problem is of order $T \approx T_{EQ1} \approx 1$ (and not, let us say, 1000, or $10^{-2}$). Combined with the contraction rate (4.34), 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 very small radial per-turn expansion multiplier $\Lambda^{(1)} \simeq 1.06$, and very strong contraction multiplier $\Lambda^{(3)} \simeq 10^{-4}$ onto the unstable manifold, figure 4.7 (a). This contraction confines, for all practical purposes, the Lorenz attractor to a 2-dimensional surface evident in the section figure 3.7.*

*In the $x_{EQ0} = (0,0,0)$ 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 4.7 (b). Thus linearization suffices to describe analytically the singular dip in the Poincaré sections of figure 3.7, and the empirical scarcity of trajectories close to $EQ_0$. (Continued in example 4.7.)*

*(E. Siminos and J. Halcrow)*

**Example 4.7** ***Lorenz flow: 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 3.7 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? As in the neighborhood of the $EQ_0$ equilibrium the dynamics is linear (see figure 4.7 (a)), 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 $EQ_0$, then resume the numerical integration.*

*What happens next? Trajectories to the left of $z$-axis shoot off along the $\mathbf{e}^{(1)}$ direction, and those to the right along $-\mathbf{e}^{(1)}$. As along the $\mathbf{e}^{(1)}$ direction $xy > 0$, the nonlinear term in the $\dot{z}$ equation (2.12) bends both branches of the $EQ_0$ unstable manifold $W^u(EQ_0)$ upwards. Then ... - never mind. Best to postpone the completion of this narrative to example 9.2, 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 9.2.)*

*(E. Siminos and J. Halcrow)*

### 4.3.2 Stability of trajectories

Next, 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):

$$e^{tA} = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k \tag{4.40}$$

$$= \lim_{m\to\infty} \left(\mathbf{1} + \frac{t}{m} A\right)^m . \tag{4.41}$$

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, as for a dynamical system the local rate of neighborhood distortion $A(x)$ depends on where we are along the trajectory. The linearized neighborhood is multiplicatively deformed along the flow, and the $m$ discrete time step approximation to $J^t$ is therefore given by a generalization of the Euler product (4.41):

$$\begin{aligned} J^t &= \lim_{m\to\infty} \prod_{n=m}^{1} (\mathbf{1} + \delta t A(x_n)) = \lim_{m\to\infty} \prod_{n=m}^{1} e^{\delta t A(x_n)} \\ &= \lim_{m\to\infty} e^{\delta t A(x_n)} e^{\delta t A(x_{m-1})} \cdots e^{\delta t A(x_2)} e^{\delta t A(x_1)} , \end{aligned} \tag{4.42}$$

where $\delta t = (t - t_0)/m$, and $x_n = x(t_0 + n\delta t)$. Slightly perverse indexing of the products indicates that in our convention the successive infinitesimal deformation are applied by multiplying from the left. The two formulas for $J^t$ agree to leading order in $\delta t$, and the $m \to \infty$ limit of this procedure is the integral

$$J^t_{ij}(x_0) = \left[ \mathbf{T} e^{\int_0^t d\tau A(x(\tau))} \right]_{ij} , \tag{4.43}$$

where $\mathbf{T}$ stands for time-ordered integration, *defined* as the continuum limit of the successive left multiplications (4.42). This integral formula for $J$ is the main conceptual result of this chapter. [exercise 4.5]

It makes evident important properties of fundamental matrices, such as that they are multiplicative along the flow,

$$J^{t+t'}(x) = J^{t'}(x')\, J^t(x), \qquad \text{where } x' = f^t(x) , \tag{4.44}$$

an immediate consequence of time-ordered product structure of (4.42). However, in practice $J$ is evaluated by integrating (4.9) along with the ODEs that define a particular flow.

in depth:
sect. 15.3, p. 263

## 4.4 Neighborhood volume

[section 15.3]
[remark 15.3]

Consider a small state space volume $\Delta V = d^d x$ centered around the point $x_0$ at time $t = 0$. The volume $\Delta V' = \Delta V(t)$ around the point $x' = x(t)$ time $t$ later is

$$\Delta V' = \frac{\Delta V'}{\Delta V} \Delta V = \left| \det \frac{\partial x'}{\partial x} \right| \Delta V = \left| \det\, J(x_0)^t \right| \Delta V , \tag{4.45}$$

so the $|\det J|$ is the ratio of the initial and the final volumes. The determinant $\det J^t(x_0) = \prod_{i=1}^{d} \Lambda_i(x_0, t)$ is the product of the multipliers. We shall refer to this

determinant as the *Jacobian* of the flow. This Jacobian is easily evaluated. Take the time derivative and use the matrix identity $\ln\det J = \mathrm{tr}\,\ln J$: [exercise 4.1]

$$\frac{d}{dt}\ln\Delta V(t) = \frac{d}{dt}\ln\det J = \mathrm{tr}\,\frac{d}{dt}\ln J = \mathrm{tr}\,\frac{1}{J}\dot{J} = \mathrm{tr}\,A = \partial_i v_i\,.$$

(Here, as elsewhere in this book, a repeated index implies summation.) As the divergence $\partial_i v_i$ is a scalar quantity, the integral in the exponent needs *no time ordering*. Integrate both sides to obtain the time evolution of an infinitesimal volume

$$\det J^t(x_0) = \exp\left[\int_0^t d\tau\,\mathrm{tr}\,\mathbf{A}(x(\tau))\right] = \exp\left[\int_0^t d\tau\,\partial_i v_i(x(\tau))\right]\,. \tag{4.46}$$

All we need to do is evaluate the time average

$$\begin{aligned}\overline{\partial_i v_i} &= \lim_{t\to\infty}\frac{1}{t}\int_0^t d\tau\,\sum_{i=1}^{d} A_{ii}(x(\tau))\\ &= \frac{1}{t}\ln\left|\prod_{i=1}^{d}\Lambda_i(x_0,t)\right| = \sum_{i=1}^{d}\lambda^{(i)}(x_0,t)\end{aligned} \tag{4.47}$$

along the trajectory. If the flow is not singular (for example, the trajectory does not run head-on into the Coulomb $1/r$ singularity), the stability matrix elements are bounded everywhere, $|A_{ij}| < M$, and so is the trace $\sum_i A_{ii}$. The time integral in (4.46) grows at most linearly with $t$, hence $\overline{\partial_i v_i}$ is bounded for all times, and numerical estimates of the $t\to\infty$ limit in (4.47) are not marred by any blowups.

Even if we were to insist on extracting $\overline{\partial_i v_i}$ from (4.42) by first multiplying fundamental matrices along the flow, and then taking the logarithm, we can avoid exponential blowups in $J^t$ by using the multiplicative structure (4.44), $\det J^{t'+t}(x_0) = \det J^{t'}(x')\det J^t(x_0)$ to restart with $J^0(x') = \mathbf{1}$ whenever the eigenvalues of $J^t(x_0)$ start getting out of hand. In numerical evaluations of Lyapunov exponents, $\lambda_i = \lim_{t\to\infty}\mu^{(i)}(x_0,t)$, the sum rule (4.47) can serve as a helpful check on the accuracy of the computation. [section 15.3]

The divergence $\partial_i v_i$ is an important characterization of the flow - it describes the behavior of a state space volume in the infinitesimal neighborhood of the trajectory. If $\partial_i v_i < 0$, 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, and the dimension of the attractor is necessarily smaller than the dimension of state space $\mathcal{M}$. If $\partial_i v_i = 0$, the flow preserves state space volume and $\det J^t = \mathbf{1}$. A flow with this property is called *incompressible*. An important class of such flows are the Hamiltonian flows considered in sect. 7.2.

But before we can get to that, Henri Roux, the perfect student always on alert, pipes up. He does not like our definition of the fundamental matrix in terms of the time-ordered exponential (4.43). Depending on the signs of multipliers, the left

hand side of (4.46) 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? 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.

## 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 *discrete* time $n$ to linear order, as in (4.5). The linearized neighborhood is transported by the fundamental matrix evaluated at a discrete set of times $n = 1, 2, \ldots$,

$$M^n_{ij}(x_0) = \left.\frac{\partial f^n_i(x)}{\partial x_j}\right|_{x=x_0} . \tag{4.48}$$

We shall refer to this Jacobian matrix also as the *monodromy* matrix, in case of periodic orbits $f^n(x) = x$. Derivative notation $M^t(x_0) \to Df^t(x_0)$ is frequently employed in the literature. As in the continuous case, we denote by $\Lambda_k$ the $k$th *eigenvalue* or multiplier of the finite time fundamental matrix $M^n(x_0)$, and $\mu^{(k)}$ the real part of $k$th *eigen-exponent*

$$\Lambda_\pm = e^{n(\mu \pm i\omega)} , \qquad |\Lambda| = e^{n\mu} .$$

For complex eigenvalue pairs the phase $\omega$ describes the rotation velocity in the plane defined by the corresponding pair of eigenvectors, with one period of rotation given by

$$T = 2\pi/\omega . \tag{4.49}$$

**Example 4.8** ***Stability of a 1-dimensional map:*** *Consider a 1-d map $f(x)$. The chain rule yields the stability of the $n$th iterate*

$$\Lambda(x_0, n) = \frac{d}{dx} f^n(x_0) = \prod_{m=0}^{n-1} f'(x_m) , \quad x_m = f^m(x_0) . \tag{4.50}$$

The 1-step product formula for the stability of the $n$th iterate of a $d$-dimensional map

$$\begin{aligned} M^n(x_0) &= M(x_{n-1}) \cdots M(x_1) M(x_0) , \\ M(x)_{kl} &= \frac{\partial}{\partial x_l} f_k(x) , \quad x_m = f^m(x_0) \end{aligned} \tag{4.51}$$

follows from the chain rule for matrix derivatives

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

If you prefer to think of a discrete time dynamics as a sequence of Poincaré section returns, then (4.51) follows from (4.44): fundamental matrices are multiplicative along the flow.

[exercise 15.1]

**Example 4.9** ***Hénon map fundamental matrix:*** *For the Hénon map (3.18) the fundamental matrix for the $n$th iterate of the map is*

$$M^n(x_0) = \prod_{m=n}^{1} \begin{pmatrix} -2ax_m & b \\ 1 & 0 \end{pmatrix}, \qquad x_m = f_1^m(x_0, y_0) \,. \tag{4.52}$$

*The determinant of the Hénon one time step fundamental matrix (4.52) is constant,*

$$\det M = \Lambda_1 \Lambda_2 = -b \tag{4.53}$$

*so 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.*

fast track:
chapter 7, p. 108

### 4.5.1 Stability of Poincaré return maps

(R. Paškauskas and P. Cvitanović)

We now relate the linear stability of the Poincaré return map $P : \mathcal{P} \to \mathcal{P}$ 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 $x' = x(\tau)$, so expanding $U(x')$ to linear order in $\delta x$ leads to the condition

$$\sum_{i=1}^{d+1} \frac{\partial U(x')}{\partial x_i} \frac{dx'_i}{dx_j}\Big|_{\mathcal{P}} = 0 \,. \tag{4.54}$$

In what follows $U_i$ 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

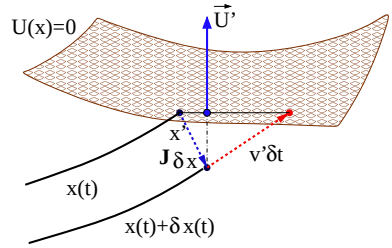



**Figure 4.8:** If $x(t)$ intersects the Poincaré section $\mathcal{P}$ at time $\tau$, the nearby $x(t) + \delta x(t)$ trajectory intersects it time $\tau + \delta t$ later. As $(U' \cdot v' \delta t) = -(U' \cdot J\,\delta x)$, the difference in arrival times is given by $\delta t = -(U' \cdot J\,\delta x)/(U' \cdot v')$.

full state space fundamental 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 fundamental matrix

$$\hat{J}(x)_{ij} = \left.\frac{dx'_i}{dx_j}\right|_{\mathcal{P}} \tag{4.55}$$

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

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

The return time variation $d\tau/dx$, figure 4.8, is eliminated by substituting this expression into the constraint (4.54),

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

yielding the projection of the full space $(d+1)$-dimensional fundamental matrix to the Poincaré map $d$-dimensional fundamental matrix:

$$\hat{J}_{ij} = \left(\delta_{ik} - \frac{v'_i\,\partial_k U'}{(v' \cdot \partial U')}\right) J_{kj}\,. \tag{4.56}$$

Substituting (4.7) we verify that the initial velocity $v(x)$ is a zero-eigenvector of $\hat{J}$

$$\hat{J} v = 0\,, \tag{4.57}$$

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

## Résumé

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 fundamental 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 is multiplication by time step fundamental matrix $M$ along the $n$ points $x_0, x_1, x_2, \ldots, x_{n-1}$ on the trajectory of $x_0$,

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

with $M(x)$ the single discrete time step fundamental matrix. In this book $\Lambda_k$ denotes the $k$th *eigenvalue* of the finite time fundamental matrix $J^t(x_0)$, and $\mu^{(k)}$ the real part of $k$th *eigen-exponent*

$$|\Lambda| = e^{n\mu} , \qquad \Lambda_\pm = e^{n(\mu \pm i\omega)} .$$

For complex eigenvalue pairs the phase $\omega$ describes rotational motion in the plane defined by the corresponding pair of eigenvectors.

The eigenvalues and eigen-directions of the fundamental matrix describe the deformation of an initial infinitesimal sphere of neighboring trajectories into an ellipsoid a finite time $t$ later. Nearby trajectories separate exponentially along unstable directions, approach each other along stable directions, and change slowly (algebraically) their distance along marginal directions. The fundamental 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 fundamental matrices are multiplicative along the flow, in dimensions higher than one their eigenvalues in general are not. This lack of multiplicativity has important repercussions for both classical and quantum dynamics.

## Commentary

**Remark 4.1** Linear flows. The subject of linear algebra generates innumerable tomes of its own; in sect. 4.2 we only sketch, and in appendix B recapitulate a few facts that our narrative relies on. They are presented at length in many textbooks. The standard references that exhaustively enumerate and explain all possible cases are Hirsch and Smale [1], and Arnol'd [1]. For ChaosBook purposes, we enjoyed the discussion in chapter 2 Meiss [2], chapter 1 of Perko [3] and chapters 3 and 5 of Glendinning [4] most.

The construction of projection operators given here is taken from refs. [6, 7]. Who wrote this down first we do not know, lineage certainly goes all the way back to Lagrange polynomials [10], but projection operators tend to get drowned in sea of algebraic details. Halmos [5] is a good early reference - but we like Harter's exposition [8, 9, 12] best, for its multitude of specific examples and physical illustrations.

The nomenclature tends to be a bit confusing. In referring to $A$ defined in (4.3) as the "stability matrix" we follow Tabor [13]. Sometimes $A$, which describes the instantaneous shear of the trajectory point $x(x_0, t)$ is referred to as the 'Jacobian matrix,' a particularly unfortunate usage when one considers linearized stability of an equilibrium point (4.30). What Jacobi had in mind in his 1841 fundamental paper [11] on the determinants today known as 'jacobians' were transformations between different coordinate frames. These are dimensionless quantities, while the dimension of $A_{ij}$ 1/[time]. More unfortunate still is referring to $J^t = e^{tA}$ as an 'evolution operator,' which here (see sect. 15.2) refers to something altogether different. In this book fundamental matrix $J^t$ always refers to (4.6), the linearized deformation after a finite time $t$, either for a continuous time flow, or a discrete time mapping.

## Exercises

4.1. **Trace-log of a matrix.** Prove that

$$\det M = e^{\mathrm{tr}\,\ln M}\,.$$

for an arbitrary nonsingular finite dimensional matrix $M$, $\det M \neq 0$.

4.2. **Stability, diagonal case.** Verify the relation (4.17)

$$J^t = e^{t\mathbf{A}} = \mathbf{U}^{-1} e^{t\mathbf{A}_D}\mathbf{U}\,, \quad \mathbf{A}_D = \mathbf{U}\mathbf{A}\mathbf{U}^{-1}\,.$$

4.3. **State space volume contraction.**

(a) Compute the Rössler flow volume contraction rate at the equilibria.

(b) 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.

(c) Compute numerically the average contraction rate (4.47) along a typical trajectory on the Rössler attractor.

(d) (optional) color-code the points on the trajectory by the sign (and perhaps the magnitude) of $\partial_i v_i - \bar{\partial}_i v_i$.

(e) Argue on basis of your results that this attractor is of dimension smaller than the state space $d = 3$.

(f) (optional) Start some trajectories on the escape side of the outer equilibrium, color-code the points on the trajectory. Is the flow volume contracting?

4.4. **Topology of the Rössler flow.** (continuation of exercise 3.1)

(a) Show that equation $|\det(A - \lambda\mathbf{1})| = 0$ for Rössler flow in the notation of exercise 2.8 can be written as

$$\lambda^3 + \lambda^2 c\,(p^{\mp} - \epsilon) + \lambda(p^{\pm}/\epsilon + 1 - c^2\epsilon p^{\mp}) \mp c\sqrt{D} = 0 \quad (4.5$$

(b) Solve (4.58) for eigenvalues $\lambda^{\pm}$ for each equilibrium as an expansion in powers of $\epsilon$. Derive

$$\begin{aligned}
\lambda_1^- &= -c + \epsilon c/(c^2+1) + o(\epsilon) \\
\lambda_2^- &= \epsilon c^3/[2(c^2+1)] + o(\epsilon^2) \\
\theta_2^- &= 1 + \epsilon/[2(c^2+1)] + o(\epsilon) \\
\lambda_1^+ &= c\epsilon(1-\epsilon) + o(\epsilon^3) \\
\lambda_2^+ &= -\epsilon^5 c^2/2 + o(\epsilon^6) \\
\theta_2^+ &= \sqrt{1 + 1/\epsilon}\,(1 + o(\epsilon))
\end{aligned} \tag{4.59}$$

Compare with exact eigenvalues. What are dynamical implications of the extravagant value of $\lambda_1^-$? (continued as exercise 12.7)

(R. Paškauskas)

4.5. **Time-ordered exponentials.** Given a time dependent matrix $V(t)$ check that the time-ordered exponential

$$\mathcal{U}(t) = \mathbf{T}e^{\int_0^t d\tau V(\tau)}$$

may be written as

$$\mathcal{U}(t) = \sum_{m=0}^{\infty} \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{m-1}} dt_m V(t_1)\cdots V(t_m)$$

and verify, by using this representation, that $\mathcal{U}(t)$ satisfies the equation

$$\dot{\mathcal{U}}(t) = V(t)\mathcal{U}(t),$$

with the initial condition $\mathcal{U}(0) = 1$.

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

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} x_n/3 \\ 2y_n \end{pmatrix} \quad y_n \le 1/2$$

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} x_n/3 + 1/2 \\ 2y_n - 1 \end{pmatrix} \quad y_n > 1/2 \,.$$

This map shrinks strips by a factor of 1/3 in the $x$-direction, and then 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?

## References


[4.1] M. W. Hirsch and S. Smale, *Differential Equations, Dynamical Systems, and Linear Algebra*, (Academic Press, San Diego 1974).

[4.2] J. D. Meiss, *Differential Dynamical Systems* (SIAM, Philadelphia 2007).

[4.3] L. Perko, *Differential Equations and Dynamical Systems* (Springer-Verlag, New York 1991).

[4.4] P. Glendinning, *Stability, Instability, and Chaos* (Cambridge Univ. Press, Cambridge 1994).

[4.5] P. R. Halmos, *Finite-dimensional vector spaces* (D. Van Nostrand, Princeton, 1958).

[4.6] P. Cvitanović, "Group theory for Feynman diagrams in non-Abelian gauge theories," *Phys. Rev. D* **14**, 1536 (1976).

[4.7] P. Cvitanović, "Classical and exceptional Lie algebras as invariance algebras," Oxford preprint 40/77 (June 1977, unpublished); available on ChaosBook.org/refs.

[4.8] W. G. Harter, *J. Math. Phys.* **10**, 4 (1969).

[4.9] 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," *Am. J. Phys.* **46**, 251 (1978).

[4.10] K. Hoffman and R. Kunze, *Linear Algebra* (Prentice-Hall, Englewood Cliffs, NJ 1971), Chapter 6.

[4.11] C. G. J. Jacobi, "De functionibus alternantibus earumque divisione per productum e differentiis elementorum conflatum," in *Collected Works*, Vol. 22, 439; *J. Reine Angew. Math. (Crelle)* (1841).

[4.12] W. G. Harter, *Principles of Symmetry, Dynamics, and Spectroscopy* (Wiley, New York 1974).

[4.13] M. Tabor, Sect 1.4 "Linear stability analysis," in *Chaos and Integrability in Nonlinear Dynamics: An Introduction* (Wiley, New York 1989), pp. 20-31.

# Chapter 5

# Cycle stability

Topological 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. Surprisingly, there 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 properties of dynamical systems independent of coordinate choice.

We now turn to the first, local class of such invariants, linear stability of periodic orbits of flows and maps. This will give us metric information about local dynamics. If you already know that the eigenvalues of periodic orbits are invariants of a flow, skip this chapter.

fast track:
chapter 7, p. 108

## 5.1 Stability of periodic orbits

As noted on page 35, a trajectory can be stationary, periodic or aperiodic. For chaotic systems almost all trajectories are aperiodic–nevertheless, equilibria and periodic orbits will turn out to be the key to unraveling chaotic dynamics. Here we note a few of the properties that makes them so precious to a theorist.

An obvious virtue of periodic orbits is that they are *topological* invariants: a fixed point remains a fixed point for any choice of coordinates, and similarly a periodic orbit remains periodic in any representation of the dynamics. Any reparametrization 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.

To prove this, we start by noting that due to the multiplicative structure (4.44) of fundamental matrices, the fundamental matrix for the $r$th repeat of a prime cycle $p$ of period $T_p$ is

$$J^{rT_p}(x) = J^{T_p}(f^{(r-1)T_p}(x)) \cdots J^{T_p}(f^{T_p}(x))J^{T_p}(x) = \left(J_p(x)\right)^r \,, \tag{5.1}$$

where $J_p(x) = J^{T_p}(x)$ is the fundamental matrix for a single traversal of the prime cycle $p$, $x \in p$ is any point on the cycle, and $f^{rT_p}(x) = x$ as $f^t(x)$ returns to $x$ every multiple of the period $T_p$. Hence, it suffices to restrict our considerations to the stability of prime cycles.

fast track:
sect. 5.2, p. 87

### 5.1.1 Nomenclature, again

When dealing with periodic orbits, some of the quantities introduced above inherit terminology from the theory of differential equations with periodic coefficients.

For instance, if we consider the equation of variations (4.2) evaluated on a periodic orbit $p$,

$$\dot{\delta x} = A(t)\delta x \,, \qquad A(t) = A(x(t)) = A(t + T_p) \,, \tag{5.2}$$

the $T_p$ periodicity of the stability matrix implies that if $\delta x(t)$ is a solution of (5.2) then also $\delta x(t + T_p)$ satisfies the same equation: moreover the two solutions are related by (see (4.6))

$$\delta x(t + T_p) = J_p(x)\delta x(t) \,. \tag{5.3}$$

Even though the fundamental matrix $J_p(x)$ depends upon $x$ (the "starting" point of the periodic orbit), its eigenvalues do not, so we may write for its eigenvectors $\mathbf{e}^{(j)}$

$$J_p(x)\mathbf{e}^{(j)}(x) = \Lambda_{p,j}\mathbf{e}^{(j)}(x) = e^{T_p(\mu_p^{(j)} + i\omega_p^{(j)})}\mathbf{e}^{(j)}(x) \,,$$

where $\mu_p^{(j)}$ and $\omega_p^{(j)}$ are independent of $x$, and expand

$$\delta x(t) = \sum_j u_j(t)\mathbf{e}^{(j)} \,.$$

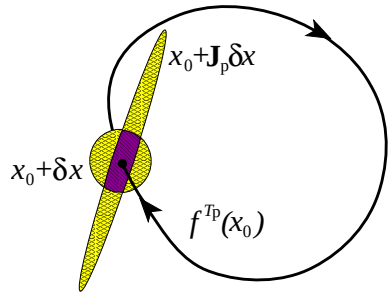



**Figure 5.1:** For a prime cycle $p$, fundamental matrix $J_p$ returns an infinitesimal spherical neighborhood of $x_0 \in p$ stretched into an ellipsoid, with overlap ratio along the expanding eigdirection $\mathbf{e}^{(i)}$ of $J_p(x)$ given by the the expanding eigenvalue $1/|\Lambda_{p,i}|$. These ratios are invariant under smooth nonlinear reparametrizations of state space coordinates, and are intrinsic property of cycle $p$.

If we take (5.3) into account, we get

$$\delta x(t+T_p) = \sum_j u_j(t+T_p)\mathbf{e}^{(j)} = \sum_j u_j(t) e^{T_p(\mu_p^{(j)}+i\omega_p^{(j)})}\mathbf{e}^{(j)}$$

which shows that the coefficients $u_j(t)$ may be written as

$$u_j(t) = e^{t(\mu_p^{(j)}+i\omega_p^{(j)})} v_j(t)$$

where $v_j(t)$ is *periodic* with period $T_p$. Thus each solution of the equation of variations may be expressed as

$$\delta x(t) = \sum_j v_j(t) e^{t(\mu_p^{(j)}+i\omega_p^{(j)})}\mathbf{e}^{(j)} \qquad v_j(t+T_p) = v_j(t)\,, \tag{5.4}$$

the form predicted by Floquet theorem for differential equations with periodic coefficients.

The continuous time $t$ appearing in (5.4) does not imply that eigenvalues of the fundamental matrix enjoy any multiplicative property: $\mu_p^{(j)}$ and $\omega_p^{(j)}$ refer to a full evolution over the complete periodic orbit. $\Lambda_{p,j}$ is called the Floquet multiplier, and $\mu_p^{(j)} + i\omega_p^{(j)}$ the Floquet or characteristic exponent, where $\Lambda_{p,j} = e^{T_p(\mu_p^{(j)}+i\omega_p^{(j)})}$.

### 5.1.2 Fundamental matrix eigenvalues and exponents

We sort the *Floquet multipliers* $\Lambda_{p,1}$, $\Lambda_{p,2}$, ..., $\Lambda_{p,d}$ of the $[d\times d]$ fundamental matrix $J_p$ evaluated on the $p$-cycle into sets $\{e, m, c\}$

$$\begin{aligned}
\text{expanding:} \quad \{\Lambda\}_e &= \{\Lambda_{p,j} : |\Lambda_{p,j}| > 1\} \\
\text{marginal:} \quad \{\Lambda\}_m &= \{\Lambda_{p,j} : |\Lambda_{p,j}| = 1\} \\
\text{contracting:} \quad \{\Lambda\}_c &= \{\Lambda_{p,j} : |\Lambda_{p,j}| < 1\}\,.
\end{aligned} \tag{5.5}$$

and denote by $\Lambda_p$ (no $j$th eigenvalue index) the product of *expanding* Floquet multipliers

$$\Lambda_p = \prod_e \Lambda_{p,e} \,. \tag{5.6}$$

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 of expanding eigenvalues $\Lambda_p$ is always real.

The stretching/contraction rates per unit time are are given by the real parts of Floquet exponents

$$\mu_p^{(i)} = \frac{1}{T_p} \ln \left|\Lambda_{p,i}\right| \,. \tag{5.7}$$

The factor $\frac{1}{T_p}$ in the definition of the Floquet exponents is motivated by its form for the linear dynamical systems, for example (4.16), as well as the fact that exponents so defined can be interpreted as Lyapunov exponents (15.33) evaluated on the prime cycle $p$. As in the three cases of (5.5), we sort the Floquet exponents $\lambda = \mu \pm i\omega$ into three sets

[section 15.3]

$$\begin{aligned}
\text{expanding:} \quad \{\lambda\}_e &= \{\lambda_p^{(i)} : \mu_p^{(i)} > 0\} \\
\text{marginal:} \quad \{\lambda\}_m &= \{\lambda_p^{(i)} : \mu_p^{(i)} = 0\} \\
\text{contracting:} \quad \{\lambda\}_c &= \{\lambda_p^{(i)} : \mu_p^{(i)} < 0\} \,.
\end{aligned} \tag{5.8}$$

A periodic orbit $p$ of a $d$-dimensional flow or a map is *stable* if real parts of all of its Floquet exponents (other than the vanishing longitudinal exponent, to be explained in sect. 5.2.1) are strictly negative, $\mu_p^{(i)} < 0$. The region of system parameter values for which a periodic orbit $p$ is stable is called the *stability window* of $p$. The set $\mathcal{M}_p$ of initial points that are asymptotically attracted to $p$ as $t \to +\infty$ (for a fixed set of system parameter values) is called the *basin of attraction* of $p$.

If *all* Floquet exponents (other than the vanishing longitudinal exponent) of *all* periodic orbits of a flow are strictly bounded away from zero, $|\mu^{(i)}| \geq \mu_{min} > 0$, the flow is said to be *hyperbolic*. Otherwise the flow is said to be *nonhyperbolic*. In particular, if all $\mu^{(i)} = 0$, the orbit is said to be *elliptic*. Such orbits proliferate in Hamiltonian flows.

[section 7.3]

We often do care about $\sigma_p^{(j)} = \Lambda_{p,j}/|\Lambda_{p,j}|$, the sign of the $j$th eigenvalue, and, if $\Lambda_{p,j}$ is complex, its phase

$$\Lambda_{p,j} = \sigma_p^{(j)} e^{\lambda_p^{(j)} T_p} = \sigma_p^{(j)} e^{(\mu_p^{(j)} \pm i\omega_p^{(j)}) T_p} \,. \tag{5.9}$$

[section 7.2]

Keeping track of this by case-by-case enumeration is an unnecessary nuisance, followed in much of the literature. To avoid this, almost all of our formulas will be stated in terms of the Floquet multipliers $\Lambda_j$ rather than in the terms of the overall signs, Floquet exponents $\lambda^{(j)}$ and phases $\omega^{(j)}$.

**Example 5.1** ***Stability of 1-$d$ map cycles:*** *The simplest example of cycle stability is afforded by 1-dimensional maps. The stability of a prime cycle $p$ follows from the chain rule (4.50) for stability of the $n_p$th iterate of the map*

$$\Lambda_p = \frac{d}{dx_0} f^{n_p}(x_0) = \prod_{m=0}^{n_p-1} f'(x_m)\,, \quad x_m = f^m(x_0)\,. \tag{5.10}$$

*$\Lambda_p$ is a property of the cycle, not the initial point, as taking any periodic point in the $p$ cycle as the initial point yields the same result.*

*A* critical point *$x_c$ is a value of $x$ for which the mapping $f(x)$ has vanishing derivative, $f'(x_c) = 0$. For future reference we note that a periodic orbit of a 1-dimensional map is* stable *if*

$$\left|\Lambda_p\right| = \left|f'(x_{n_p})f'(x_{n_p-1})\cdots f'(x_2)f'(x_1)\right| < 1\,,$$

*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 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 $\left|\Lambda_p\right| > 1$, $p$-cycle is* unstable.

**Example 5.2** ***Stability of cycles for maps:*** *No matter what method we use to determine the unstable cycles, the theory to be developed here requires that their Floquet multipliers be evaluated as well. For maps a fundamental matrix is easily evaluated by picking any cycle point as a starting point, running once around a prime cycle, and multiplying the individual cycle point fundamental matrices according to (4.51). For example, the fundamental matrix $M_p$ for a Hénon map (3.18) prime cycle $p$ of length $n_p$ is given by (4.52),*

$$M_p(x_0) = \prod_{k=n_p}^{1} \begin{pmatrix} -2ax_k & b \\ 1 & 0 \end{pmatrix}, \qquad x_k \in p\,,$$

*and the fundamental matrix $M_p$ for a 2-dimensional billiard prime cycle $p$ of length $n_p$*

$$M_p = (-1)^{n_p} \prod_{k=n_p}^{1} \begin{pmatrix} 1 & \tau_k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ r_k & 1 \end{pmatrix}$$

*follows from (8.11) of chapter 8. We shall compute Floquet multipliers of Hénon map cycles once we learn how to find their periodic orbits, see exercise 12.10.*

## 5.2 Cycle Floquet multipliers are cycle invariants

The 1-dimensional map cycle Floquet multiplier $\Lambda_p$ 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 fundamental matrix $J_p(x_0)$ in (5.1) does depend on the choice of coordinates and the initial point $x_0 \in p$. Nevertheless, as we shall now show, the cycle *Floquet multipliers* are intrinsic property of a cycle also for multi-dimensional flows. Consider the

$i$th eigenvalue, eigenvector pair $(\Lambda_{p,i}, \mathbf{e}^{(i)})$ computed from $J_p$ evaluated at a cycle point,

$$J_p(x)\mathbf{e}^{(i)}(x) = \Lambda_{p,i}\mathbf{e}^{(i)}(x)\,, \quad x \in p\,. \tag{5.11}$$

Consider another point on the cycle at time $t$ later, $x' = f^t(x)$ whose fundamental matrix is $J_p(x')$. By the group property (4.44), $J^{T_p+t} = J^{t+T_p}$, and the fundamental matrix at $x'$ can be written either as

$$J^{T_p+t}(x) = J^{T_p}(x')J^t(x) = J_p(x')J^t(x)\,, \quad \text{or} \quad J_p(x')J^t(x) = J^t(x)J_p(x)\,.$$

Multiplying (5.11) by $J^t(x)$, we find that the fundamental matrix evaluated at $x'$ has the same eigenvalue,

$$J_p(x')\mathbf{e}^{(i)}(x') = \Lambda_{p,i}\mathbf{e}^{(i)}(x')\,, \quad \mathbf{e}^{(i)}(x') = J^t(x)\mathbf{e}^{(i)}(x)\,, \tag{5.12}$$

but with the eigenvector $\mathbf{e}^{(i)}$ transported along the flow $x \to x'$ to $\mathbf{e}^{(i)}(x') = J^t(x)\mathbf{e}^{(i)}(x)$. Hence, $J_p$ evaluated anywhere along the cycle has the same set of Floquet multipliers $\{\Lambda_{p,1}, \Lambda_{p,2}, \cdots \Lambda_{p,d-1}, 1\}$. As quantities such as tr $J_p(x)$, det $J_p(x)$ depend only on the eigenvalues of $J_p(x)$ and not on the starting point $x$, in expressions such as $\det\left(\mathbf{1} - J_p^r(x)\right)$ we may omit reference to $x$:

$$\det\left(\mathbf{1} - J_p^r\right) = \det\left(\mathbf{1} - J_p^r(x)\right) \quad \text{for any } x \in p\,. \tag{5.13}$$

We postpone the proof that the cycle Floquet multipliers are smooth conjugacy invariants of the flow to sect. 6.6.

### 5.2.1 Marginal eigenvalues

The presence of marginal eigenvalues signals either a continuous symmetry of the flow (which one should immediately exploit to simplify the problem), or a non-hyperbolicity 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, deal with Jordan type subspaces (see example 4.3), and sub-exponential growth rates, such as $t^\alpha$.

[chapter 23]

[exercise 5.1]

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 eigenvalue*:

As $J^t(x)$ transports the velocity field $v(x)$ by (4.7), after a complete period

$$J_p(x)v(x) = v(x)\,, \tag{5.14}$$

so a periodic orbit of a *flow* always has an eigenvector $\mathbf{e}^{(d)}(x) = v(x)$ parallel to the local velocity field with the unit eigenvalue

$$\Lambda_{p,d} = 1\,, \qquad \lambda_p^{(d)} = 0\,. \tag{5.15}$$

[exercise 6.2]

The continuous invariance that gives rise to this marginal eigenvalue is the invariance of a cycle under a translation of its points along the cycle: two points on the cycle (see figure 4.3) initially distance $\delta x$ apart, $x'(0) - x(0) = \delta x(0)$, are separated by the exactly same $\delta x$ after a full period $T_p$. As we shall see in sect. 5.3, this marginal stability direction can be eliminated by cutting the cycle by a Poincaré section and eliminating the continuous flow fundamental matrix in favor of the fundamental matrix of the Poincaré return map.

If the flow is governed by a time-independent Hamiltonian, the energy is conserved, and that leads to an additional marginal eigenvalue (remember, by symplectic invariance (7.19) real eigenvalues come in pairs).

## 5.3 Stability of Poincaré 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 Poincaré return map $P$ with stability (4.56)

$$\hat{J}_{ij} = \left(\delta_{ik} - \frac{v_i\, U_k}{(v \cdot U)}\right) J_{kj}\,, \tag{5.16}$$

with all primes dropped, as the initial and the final points coincide, $x' = f^{T_p}(x) = x$. 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 (4.57) that the velocity $v(x)$ is a zero-eigenvector of the Poincaré section fundamental matrix, $\hat{J}v = 0$. Consider next $(\Lambda_{p,\alpha}, \mathbf{e}^{(\alpha)})$, the full state space $\alpha$th (eigenvalue, eigenvector) pair (5.11), evaluated at a cycle point on a Poincaré section,

$$J(x)\mathbf{e}^{(\alpha)}(x) = \Lambda_\alpha \mathbf{e}^{(\alpha)}(x)\,, \quad x \in \mathcal{P}\,. \tag{5.17}$$

Multiplying (5.16) by $\mathbf{e}^{(\alpha)}$ and inserting (5.17), we find that the full state space fundamental matrix and the Poincaré section fundamental matrix $\hat{J}$ has the same eigenvalue

$$\hat{J}(x)\hat{\mathbf{e}}^{(\alpha)}(x) = \Lambda_\alpha \hat{\mathbf{e}}^{(\alpha)}(x)\,, \quad x \in \mathcal{P}\,, \tag{5.18}$$

where $\hat{\mathbf{e}}^{(\alpha)}$ is a projection of the full state space eigenvector onto the Poincaré section:

$$(\hat{\mathbf{e}}^{(\alpha)})_i = \left(\delta_{ik} - \frac{v_i\ U_k}{(v \cdot U)}\right)(\mathbf{e}^{(\alpha)})_k\,. \tag{5.19}$$

Hence, $\hat{J}_p$ evaluated on any Poincaré section point along the cycle $p$ has the same set of Floquet multipliers $\{\Lambda_{p,1}, \Lambda_{p,2}, \cdots \Lambda_{p,d}\}$ as the full state space fundamental matrix $J_p$.

As established in (4.57), due to the continuous symmetry (time invariance) $\hat{J}_p$ is a rank $d-1$ matrix. We shall refer to any such full rank $[(d-N)\times(d-N)]$ submatrix with $N$ 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).

## 5.4 There goes the neighborhood

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. The sub-volume $|\mathcal{M}_i| = \prod_i^e \Delta x_i$ 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 $\propto 1/|\Lambda_p|$ where $\Lambda_p$ is the product of expanding eigenvalues (5.6) only; contracting ones play a secondary role. So secondary that even infinitely many of them will not matter.

So the physically important information is carried by the expanding sub-volume, not the total volume computed so easily in (4.47). That is also the reason why the dissipative and the Hamiltonian chaotic flows are much more alike than one would have naively expected 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 form an

infinite set of *metric invariants*: The Floquet multipliers of a periodic orbit remain invariant under any smooth nonlinear change of coordinates $f \to h \circ f \circ h^{-1}$ .

We shall show in chapter 10 that extending their local stability eigendirections into stable and unstable manifolds yields important global information about the topological organization of state space.

In hyperbolic systems what matters 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 eigenvalues of $J_p$. As long as the number of unstable directions is finite, the theory can be applied to flows of arbitrarily high dimension.

## Commentary

**Remark 5.1** Floquet theory. Floquet theory is a classical subject in the theory of differential equations [2]. 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 [3], or quasimomenta in the quantum theory of time-periodic Hamiltonians.

## Exercises

5.1. **A limit cycle with analytic Floquet exponent.** There are only two examples of nonlinear flows for which the stability eigenvalues can be evaluated analytically. Both are cheats. One example is the 2-*d* 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 stability eigenvalues 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 stability eigenvalues can be evaluated analytically.

## References

[5.1] J. Moehlis and K. Josić, "Periodic Orbit," www.scholarpedia.org/article/Periodic_Orbit.

[5.2] G. Floquet, "Sur les equations differentielles lineaires a coefficients periodique," *Ann. Ecole Norm. Ser. 2,* **12**, 47 (1883); E.L. Ince, *Ordinary Differential Equations* (Dover, New York 1953).

[5.3] N.W. Ashcroft and N.D. Mermin, *Solid State Physics* (Holt, Rinehart and Winston, New York 1976).

# Chapter 6

# Get straight

> We owe it to a book to withhold judgment until we reach page 100.
>
> —Henrietta McNutt, George Johnson's seventh-grade English teacher

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 too much to hope for. Still, as we shall now show, we can make some headway on straightening out the flow locally.

Even though such nonlinear coordinate transformations are very important, especially in celestial mechanics, we shall not necessarily use them much in what follows, so you can safely skip this chapter on the first reading. Except, perhaps, you might want to convince yourself that cycle stabilities are indeed metric invariants of flows (sect. 6.6), and you might like transformations that turn a Keplerian ellipse into a harmonic oscillator (example 6.2) and regularize the 2-body Coulomb collisions (sect. 6.3) in classical helium.

fast track:
chapter 14, p. 235

## 6.1 Changing coordinates

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 vector

field will also change. The vector field lives in a (hyper)plane tangent to state space and changing the coordinates of state space 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 $\mathcal{M}' = h(\mathcal{M})$, with a point $x \in \mathcal{M}$ related to a point $y \in \mathcal{M}'$ by

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

The change of coordinates must be one-to-one and span both $\mathcal{M}$ and $\mathcal{M}'$, 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*.

The evolution rule $g^t(y_0)$ on $\mathcal{M}'$ can be computed from the evolution rule $f^t(x_0)$ on $\mathcal{M}$ by taking the initial point $y_0 \in \mathcal{M}'$, going back to $\mathcal{M}$, evolving, and then mapping the final point $x(t)$ back to $\mathcal{M}'$:

$$y(t) = g^t(y_0) = h \circ f^t \circ h^{-1}(y_0) \,. \tag{6.1}$$

Here '$\circ$' stands for functional composition $h \circ f(x) = h(f(x))$, so (6.1) is a shorthand for $y(t) = h(f^t(h^{-1}(y_0)))$.

The vector field $\dot{x} = v(x)$ in $\mathcal{M}$, locally tangent to the flow $f^t$, is related to the flow by differentiation (2.5) along the trajectory. The vector field $\dot{y} = w(y)$ in $\mathcal{M}'$, locally tangent to $g^t$ follows by the chain rule:

[exercise 6.1]

$$\begin{aligned} w(y) &= \left.\frac{dg^t}{dt}(y)\right|_{t=0} = \left.\frac{d}{dt}\left(h \circ f^t \circ h^{-1}(y)\right)\right|_{t=0} \\ &= h'(h^{-1}(y))\, v(h^{-1}(y)) = h'(x)\, v(x) \,. \end{aligned} \tag{6.2}$$

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

$$\left.\frac{\partial h}{\partial x}\right|_x \cdot \left.\frac{\partial h^{-1}}{\partial y}\right|_y = 1 \quad \rightarrow \quad \frac{\partial h}{\partial x}(x) = \left[\frac{\partial h^{-1}}{\partial y}(y)\right]^{-1} \,, \tag{6.3}$$

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

$$\dot{y}_i = w_i(y) = \left[\frac{\partial h^{-1}}{\partial y}(y)\right]^{-1}_{ij} v_j(h^{-1}(y)) \,. \tag{6.4}$$

Imagine that the state space is 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 $\mathcal{M}'$, 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 Jacobian matrix $\partial_j h_i$, hence the simple transformation law (6.2) for the velocity fields.

The time itself is a parametrization of points along flow lines, and it can also be reparameterized, $s = s(t)$, with the attendant modification of (6.4). An example is the 2-body collision regularization of the helium Hamiltonian (7.6), to be undertaken in sect. 6.3 below.

## 6.2 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. In general and globally these are just words, as we have no clue how to pick such 'canonical' representative, 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 assures us that there exists a solution (at least for a short time interval) and what the solution 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 (2.8), and points at which $v$ is infinite. 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

$$\begin{aligned} &\dot{y}_1 = \dot{y}_2 = \cdots = \dot{y}_{d-1} = 0 \\ &\dot{y}_d = 1\,, \end{aligned} \tag{6.5}$$

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 finite time $\tau$ action

$$\begin{aligned} y'_i &= y_i\,, \qquad i = 1, 2, \ldots, d-1 \\ y'_d &= y_d + \tau\,. \end{aligned}$$

[exercise 9.7]

**Example 6.1** ***Harmonic oscillator, rectified:*** *As a simple example of global rectification of a flow consider the harmonic oscillator*

$$\dot{q} = p\,, \qquad \dot{p} = -q\,. \tag{6.6}$$

*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 $y = (r, \theta)$:*

$$h^{-1} : \left\{ \begin{array}{lll} q &=& h_1^{-1}(r,\theta) = r\cos\theta \\ p &=& h_2^{-1}(r,\theta) = r\sin\theta \end{array} \right. . \tag{6.7}$$

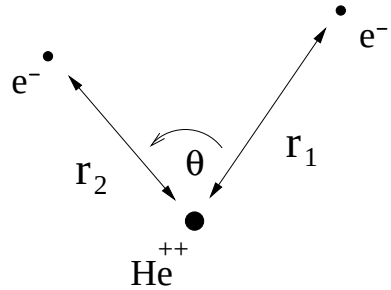



**Figure 6.1:** Coordinates for the helium three body problem in the plane.

*The Jacobian matrix of the transformation is*

$$h' = \begin{pmatrix} \cos\theta & \sin\theta \\ -\dfrac{\sin\theta}{r} & \dfrac{\cos\theta}{r} \end{pmatrix} \tag{6.8}$$

*resulting in (6.4) of rectified form*

[exercise 5.1]

$$\begin{pmatrix} \dot{r} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\dfrac{\sin\theta}{r} & \dfrac{\cos\theta}{r} \end{pmatrix} \begin{pmatrix} \dot{q} \\ \dot{p} \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}. \tag{6.9}$$

*In the new coordinates the radial coordinate $r$ is constant, and the angular coordinate $\theta$ 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 had 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 $r = 0$.*

## 6.3 Classical dynamics of 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 classical and quantum dynamics of the full three-body problem.

The classical 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 $+2e$.

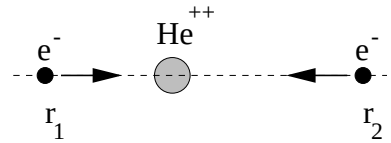



**Figure 6.2:** Collinear helium, with the two electrons on opposite sides of the nucleus.

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 $\Theta$, see figure 6.1.

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 6.2, then there is no force acting on the electrons perpendicular to the common interparticle axis. That is, if we start the classical system on the dynamical subspace $\Theta = \pi, \frac{d}{dt}\Theta = 0$, the three particles will remain in this *collinear configuration* for all times.

### 6.3.1 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

$$H = \frac{1}{2m_e}\left(p_1^2 + p_2^2\right) - \frac{2e^2}{r_1} - \frac{2e^2}{r_2} + \frac{e^2}{r_1 + r_2} = E\,, \tag{6.10}$$

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 $(r_1, r_2, p_1, p_2) \to (r_1(E), r_2(E), p_1(E), p_2(E))$ .

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 $r_i \to \infty$, $i = 1, 2$. 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

$$H = \frac{p_1^2}{2} + \frac{p_2^2}{2} - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_1 + r_2} = -1\,. \tag{6.11}$$

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.

### 6.3.2 Regularization of two–body collisions

Next, we have a closer look at the singularities in the Hamiltonian (6.11). 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 classical 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

$$H = \frac{1}{2}p^2 - \frac{2}{x} = E\,. \tag{6.12}$$

**Example 6.2** ***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*

$$\sqrt{x}dx = \sqrt{2}dt\,, \qquad x = (3t)^{\frac{2}{3}}\,, \tag{6.13}$$

*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 (6.12) we choose the time transformation $dt = xd\tau$ which lifts the $|x| \to 0$ singularity in (6.12) and leads to a new Hamiltonian*

$$\mathcal{H} = \frac{1}{2}xp^2 - 2 - Ex = 0. \tag{6.14}$$

*The solution (6.13) is now parameterized by the fictitous time $d\tau$ through a pair of equations*

$$x = \tau^2\,, \qquad t = \frac{1}{3}\tau^3\,.$$

*The equations of motion are, however, still singular as $x \to 0$:*

$$\frac{d^2x}{d\tau^2} = -\frac{1}{2x}\frac{dx}{d\tau} + xE\,.$$

*Appearance of the square root in (6.13) now suggests a canonical transformation of form*

$$x = Q^2\,, \quad p = \frac{P}{2Q} \tag{6.15}$$

*which maps the Kepler problem into that of a harmonic oscillator with Hamiltonian*

$$H(Q,P) = \frac{1}{8}P^2 - EQ^2 = 2, \tag{6.16}$$

*with all singularities completely removed.*

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 (6.15), by introducing a new vector $Q$ with property $r = |Q|^2$. In this simple 1-dimensional example the KS transformation can be implemented by

$$r_1 = Q_1^2\,, \qquad r_2 = Q_2^2\,, \qquad p_1 = \frac{P_1}{2Q_1}\,, \qquad p_2 = \frac{P_2}{2Q_2} \tag{6.17}$$

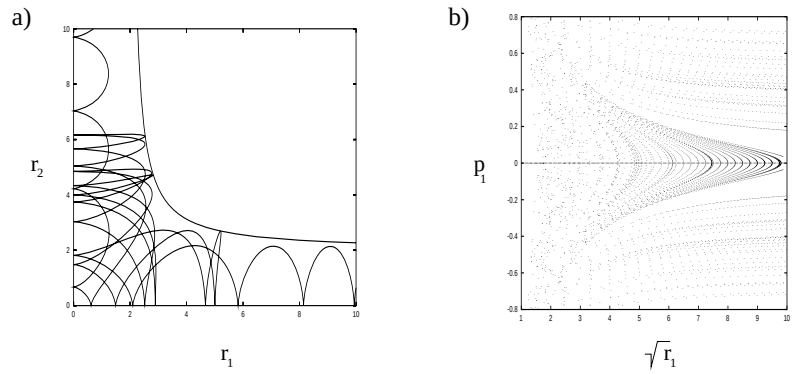



**Figure 6.3:** (a) A typical trajectory in the $[r_1, r_2]$ plane; the trajectory enters here along the $r_1$ axis and escapes to infinity along the $r_2$ axis; (b) Poincaré map ($r_2$=0) for collinear helium. Strong chaos prevails for small $r_1$ near the nucleus.

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 $\tau$ is conserved, if we consider the Hamiltonian

$$H_{ko} = \frac{1}{8}(Q_2^2 P_1^2 + Q_1^2 P_2^2) - 2R_{12}^2 + Q_1^2 Q_2^2(-E + 1/R_{12}^2) = 0 \tag{6.18}$$

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

$$\begin{aligned} \dot{P}_1 &= 2Q_1\left[2 - \frac{P_2^2}{8} - Q_2^2\left(1 + \frac{Q_2^2}{R_{12}^4}\right)\right]; \qquad \dot{Q}_1 = \frac{1}{4}P_1 Q_2^2 \\ \dot{P}_2 &= 2Q_2\left[2 - \frac{P_1^2}{8} - Q_1^2\left(1 + \frac{Q_1^2}{R_{12}^4}\right)\right]; \qquad \dot{Q}_2 = \frac{1}{4}P_2 Q_1^2. \end{aligned} \tag{6.19}$$

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 (6.19) 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 (6.18) 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 fast-forward to quantum chaos part of `ChaosBook.org`; we are already 'almost' ready to quantize helium by semiclassical methods.

fast track:
chapter 5, p. 83

## 6.4 Rectification of maps

In sect. 6.2 we had 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 are needed in 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.6.5 we will extend the method to periodic orbits.

### 6.4.1 Rectification of a fixed point in one dimension

[exercise 6.2]

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 term-by-term 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

$$f(x) = h^{-1}(\Lambda h(x)) \tag{6.20}$$

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

$$\begin{aligned} f(x) &= \Lambda x + x^2 f_2 + x^3 f_3 + \dots \\ h^{-1}(x) &= x + x^2 h_2 + x^3 h_3 + \dots \quad . \end{aligned} \tag{6.21}$$

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 (6.20); 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 first $k-1$ derivatives vanishing, the conjugacy is to the $k$th *normal form*.

In several dimensions, $\Lambda$ 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 (5.7).

[remark 6.3]

## 6.5 Rectification of a 1-dimensional periodic orbit

In sect. 6.4.1 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 cycle point $x_a$. We expand the map $f$ around each cycle point along the cycle,

$$y_a(\phi) = f_a(\phi) - x_{a+1} = \phi f_{a,1} + \phi^2 f_{a,2} + \dots \tag{6.22}$$

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 (6.21).

For a periodic orbit the conjugation formula (6.20) generalizes to

$$f_a(\phi) = h_{a+1}^{-1}(f_a'(0)h_a(\phi)), \qquad a = 1, 2, \cdots, n,$$

point by point. The conjugationg functions $h_a$ are obtained in the same way as before, by equating coefficients of the expansion (6.21), 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,

$$f^n(x_a + \phi) = h_a^{-1}(\Lambda h_a(\phi)) + x_a . \tag{6.23}$$

evaluated at each cycle point $a$. Again we have the freedom to set $h_a'(0) = 1$ for all $a$. [remark 6.2]

### 6.5.1 Repeats of cycles

We have traded in our initial nonlinear map $f$ for a (locally) linear map $\Lambda y$ 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 repeats of it are trivialized; from the conjugation formula (6.21) 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 arbitrary repeats; the answer will depend on the conjugacy function $h(x)$ computed for a *single* repeat, and all the dependence on the repeat number will be carried by factors polynomial in $\Lambda^r$, 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, be it in celestial mechanics (where the method was born), be it the quantum or stochastic corrections to 'semiclassical' approximations.

## 6.6 Cycle Floquet multipliers are metric invariants

In sect. 5.2 we have established that for a given flow the cycle Floquet multipliers are intrinsic to a given cycle, independent of the starting point along the cycle. Now we can 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.

That the cycle Floquet multipliers are an invariant property of the given dynamical system follows from elementary considerations of sect. 6.1: 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 (6.4), a reparameterization and a coordinate transformation $g = h \circ f \circ h^{-1}$. 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; $h$ must be a *smooth conjugacy* which maps nearby cycle points of $f$ into nearby cycle points of $g$. This smoothness guarantees that the cycles are not only topological invariants, but that their linearized neighborhoods are also metrically invariant. For a fixed point $f(x) = x$ of a 1-dimensional map this follows from the chain rule for derivatives,

$$\begin{aligned} g'(y) &= h'(f \circ h^{-1}(y)) f'(h^{-1}(y)) \frac{1}{h'(x)} \\ &= h'(x) f'(x) \frac{1}{h'(x)} = f'(x) \,. \end{aligned} \tag{6.24}$$

In $d$ dimensions the relationship between the maps in different coordinate representations is again $g \circ h = h \circ f$. We now make the matrix structure of relation (6.3) explicit:

$$\Gamma_{ik}(x) = \left.\frac{\partial h_i}{\partial x_k}\right|_x \qquad \text{and} \qquad \Gamma^{-1}_{ik}(x) = \left.\frac{\partial h^{-1}_i}{\partial y_k}\right|_{h(x)} ,$$

i.e., $\Gamma_{ik}(x)$ is the matrix inverse of $\Gamma^{-1}_{ik}(x)$. The chain rule now relates $M'$, the the fundamental matrix of the map $g$ to the fundamental matrix of map $f$:

$$M'_{ij}(h(x)) = \Gamma_{ik}(f(x)) M_{kl}(x) \Gamma^{-1}_{lj}(x) \,. \tag{6.25}$$

If $x$ is a fixed point then (6.25) is a *similarity* transformation and thus preserves eigenvalues: it is easy to verify that in the case of period $n_p$ cycle again $M'^p(h(x))$ and $M^p(x)$ are related by a similarity transformation (note that this is not true for $M^r(x)$ with $r \neq n_p$). As stability of a flow can always be reduced to stability of a Poincaré section 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.

[exercise 6.2]

in depth:
appendix B.3, p. 659

## Résumé

Dynamics $(\mathcal{M}, f)$ is invariant under the group of all smooth conjugacies

$$(\mathcal{M}, f) \;\to\; (\mathcal{M}', g) = (h(\mathcal{M}), h \circ f \circ h^{-1})\,.$$

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 $\mathcal{M} \to h(\mathcal{M})$ smooth conjugacies.

The $2D$-dimensional phase space of an integrable Hamiltonian system of $D$ degrees of freedom is fully foliated by $D$-tori. In the same spirit, for a uniformly hyperbolic, chaotic dynamical system one would like to change into a coordinate frame where the stable/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 developping invariant notions of what a given dynamical system is. The totality of smooth one-to-one nonlinear coordinate transformations $h$ which 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, such as the rotation group $O(d)$ or the Galilean group of all rotations and translations in $\mathbb{R}^d$. 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 stability eigenvalues will turn out to be a good start.

## Commentary

**Remark 6.1** Rectification of flows. See Section 2.2.5 of ref. [12] 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 6.2** Rectification of maps. The methods outlined above are standard in the analysis of fixed points and construction of normal forms for bifurcations, see for example ref. [22, 2, 4, 5, 6, 7, 8, 9, 9]. The geometry underlying such methods is pretty, and we enjoyed reading, for example, Percival and Richards [10], chaps. 2 and 4 of Ozorio de Almeida's monograph [11], and, as always, Arnol'd [1].

Recursive formulas for evaluation of derivatives needed to evaluate (6.21) are given, for example, in Appendix A of ref. [5]. Section 10.6 of Ref. [13] describes in detail the smooth conjugacy that relates the Ulam map to the tent map.

**Remark 6.3** 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 $\Lambda_{p,1}, \Lambda_{p,2}, \ldots, \Lambda_{p,d}$ are the Floquet multipliers of the fundamental matrix, then they are in resonance if there exist integers $n_1, \ldots, n_d$ such that

$$(\Lambda_{p,1})^{n_1}(\Lambda_{p,2})^{n_2} \cdots (\Lambda_{p,d})^{n_d} = 1 .$$

If there is resonance, then one may get corrections to the basic conjugation formulas in the form of monomials in the variables of the map. (R. Mainieri)

## Exercises

6.1. **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 the $\mathcal{M}$ space 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?

6.2. **Linearization for maps.** Let $f : C \to C$ be a map from the complex numbers into themselves, with a fixed point at the origin and analytic there. By manipulating power series, find the first few terms of the map $h$ that conjugates $f$ to $\alpha z$, 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. Can you formulate these conditions by examining the series?

(difficulty: medium) (R. Mainieri)

6.3. **Ulam and tent maps.** Show that the smooth conjugacy (6.1)

$$\begin{aligned} g(y_0) &= h \circ f \circ h^{-1}(y_0) \\ y &= h(x) = \sin^2(\pi x/2), \end{aligned}$$

conjugates the tent map $f(x) = 1 - 2|x - 1/2|$ into the Ulam map $g(y) = 4y(1 - y)$. (Continued as exercise 12.1.)

## References


[6.1] V.I. Arnol'd, *Ordinary Differential Equations* (Springer-Verlag, New York 1992).

[6.2] C. Simo, "On the analytical and numerical approximation of invariant manifolds," in D. Baenest and C. Froeschlé, Les Méthodes Modernes de la Mécanique Céleste (Goutelas 1989), p. 285.

[6.3] C. Simo, in *Dynamics and Mission Design Near Libration Points*, Vol. 1-4, (World Sci. Pub., Monograph Ser. Math., 2000-2001).

[6.4] C. L. Siegel. Iteration of analytic functions. *Ann. Math.*, 43:607–612, 1942.

[6.5] J. Moser. *Ann. Scuola Norm. Super. Pisa*, 20:265–315, 1966; 20:499–535, 1966.

[6.6] S. Sternberg. *Amer. J. Math.*, 79:809, 1957; 80:623, 1958; 81:578, 1959.

[6.7] K.-T. Chen. *Amer. J. Math.*, 85:693–722, 1963.

[6.8] G.R. Belitskiĭ. *Russian Math. Surveys*, 31:107–177, 1978.

[6.9] A.D. Brjuno. *Trans. Moscow Math. Soc.*, 25:131–288, 1971; 26:199–238, 1972.

[6.10] I. Percival and D. Richards, *Introduction to Dynamics* (Cambridge Univ. Press, Cambridge, 1982).

[6.11] A.M. Ozorio de Almeida, *Hamiltonian Systems: Chaos and Quantization* (Cambridge University Press, Cambridge, 1988).

[6.12] G. W. Bluman and S. Kumei, *Symmetries and Differential Equations* (Springer, New York 1989).

[6.13] H.-O. Peitgen, H. Jürgens and D. Saupe, *Chaos and Fractals* (Springer-Verlag, Berlin 1992).

# Chapter 7

# Hamiltonian dynamics

Truth is rarely pure, and never simple.
—Oscar Wilde

You might think that the strangeness of contracting flows, flows such as the Rössler flow of figure 2.5 is of concern only to chemists; real physicists do Hamiltonian dynamics, right? Now, that's full of chaos, too! While it is easier to visualize aperiodic dynamics when a flow is contracting onto a lower-dimensional attracting set, there are plenty examples of chaotic flows that do preserve the full symplectic invariance of Hamiltonian dynamics. The whole story started in fact with Poincaré's restricted 3-body problem, a realization that chaos rules also in general (non-Hamiltonian) flows came much later.

Here we briefly review parts of classical dynamics that we will need later on; symplectic invariance, canonical transformations, and stability of Hamiltonian flows. We discuss billiard dynamics in some detail in chapter 8.

## 7.1 Hamiltonian flows

(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 B]

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i}, \tag{7.1}$$

with the $2D$ phase space coordinates $x$ split into the configuration space coordinates and the conjugate momenta of a Hamiltonian system with $D$ degrees of freedom (dof):

$$x = (\mathbf{q}, \mathbf{p}), \qquad \mathbf{q} = (q_1, q_2, \ldots, q_D), \qquad \mathbf{p} = (p_1, p_2, \ldots, p_D). \tag{7.2}$$

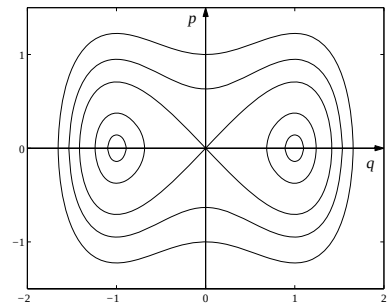



**Figure 7.1:** Phase plane of the unforced, undamped Duffing oscillator. The trajectories lie on level sets of the Hamiltonian (7.4).

The energy, or the value of the Hamiltonian function at the state space point $x = (\mathbf{q}, \mathbf{p})$ is constant along the trajectory $x(t)$,

$$\begin{aligned} \frac{d}{dt} H(\mathbf{q}(t), \mathbf{p}(t)) &= \frac{\partial H}{\partial q_i} \dot{q}_i(t) + \frac{\partial H}{\partial p_i} \dot{p}_i(t) \\ &= \frac{\partial H}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i} \frac{\partial H}{\partial q_i} = 0 \,, \end{aligned} \tag{7.3}$$

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.

**Example 7.1** ***Unforced undamped Duffing oscillator:*** *When the damping term is removed from the Duffing oscillator (2.7), the system can be written in Hamiltonian form with the Hamiltonian*

$$H(q, p) = \frac{p^2}{2} - \frac{q^2}{2} + \frac{q^4}{4}. \tag{7.4}$$

*This is a 1-dof Hamiltonian system, with a 2-dimensional state space, the plane* $(q, p)$*. The Hamilton's equations (7.1) are*

$$\dot{q} = p \,, \qquad \dot{p} = q - q^3 \,. \tag{7.5}$$

*For 1-dof systems, the 'surfaces' of constant energy (7.3) are simply curves in the phase plane* $(q, p)$*, and the dynamics is very simple: the curves of constant energy* are *the trajectories, as shown in figure 7.1.*

Thus all 1-dof systems are *integrable*, in the sense that the entire phase plane is foliated by curves of constant energy, 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 6.1]

**Example 7.2** ***Collinear helium:*** *In the quantum chaos part of* `ChaosBook.org` *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*

$$H = \frac{1}{2} p_1^2 + \frac{1}{2} p_2^2 - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_1 + r_2} \,. \tag{7.6}$$

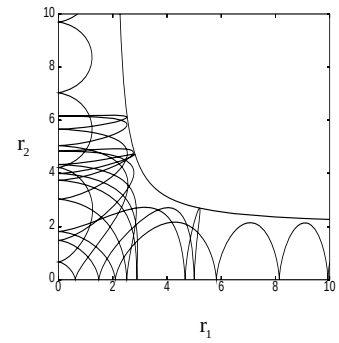



**Figure 7.2:** A typical collinear helium trajectory in the $[r_1, r_2]$ plane; the trajectory enters along the $r_1$-axis and then, like almost every other trajectory, after a few bounces escapes to infinity, in this case along the $r_2$-axis.

*Collinear helium has 2 dof, and thus a 4-dimensional phase space $\mathcal{M}$, which energy conservation reduces to 3 dimensions. The dynamics can be projected onto the 2-dimensional configuration plane, the $(r_1, r_2)$, $r_i \geq 0$ quadrant, figure 7.2. 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.*

Note an important property of Hamiltonian flows: if the Hamilton equations (7.1) 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. The symplectic invariance requirements are actually more stringent than just the phase space volume conservation, as we shall see in the next section.

## 7.2 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 stability eigenvalues by a factor of 2 or 4.

### 7.2.1 Canonical transformations

The equations of motion for a time-independent, $D$-dof Hamiltonian (7.1) can be written

$$\dot{x}_i = \omega_{ij} H_j(x)\,, \qquad \omega = \begin{pmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{pmatrix}, \quad H_j(x) = \frac{\partial}{\partial x_j} H(x)\,, \tag{7.7}$$

where $x = (\mathbf{q}, \mathbf{p}) \in \mathcal{M}$ is a phase space point, $H_k = \partial_k H$ is the column vector of partial derivatives of $H$, $\mathbf{I}$ is the $[D{\times}D]$ unit matrix, and $\omega$ the $[2D{\times}2D]$ *symplectic* form

$$\omega^T = -\omega\,, \qquad \omega^2 = -\mathbf{1}\,. \tag{7.8}$$

The evolution of $J^t$ (4.6) is again determined by the stability matrix $A$, (4.9):

$$\frac{d}{dt}J^t(x) = A(x)J^t(x)\,, \qquad A_{ij}(x) = \omega_{ik}\,H_{kj}(x)\,, \tag{7.9}$$

where the matrix of second derivatives $H_{kn} = \partial_k\partial_n H$ is called the *Hessian matrix*. From the symmetry of $H_{kn}$ it follows that

$$A^T\omega + \omega A = 0\,. \tag{7.10}$$

This is the defining property for infinitesimal generators of *symplectic* (or canonical) transformations, transformations which leave the symplectic form $\omega$ invariant.

Symplectic matrices are by definition linear transformations that leave the (antisymmetric) quadratic form $x_i\omega_{ij}y_j$ invariant. This immediately implies that any symplectic matrix satisfies

$$Q^T\omega Q \;=\; \omega\,, \tag{7.11}$$

and – when $Q$ is close to the identity $Q = \mathbf{1} + \delta t A$ – it follows that that $A$ must satisfy (7.10).

Consider now a smooth nonlinear change of variables of form $y_i = h_i(x)$, and define a new function $K(x) = H(h(x))$. Under which conditions does $K$ generate a Hamiltonian flow? In what follows we will use the notation $\tilde{\partial}_j = \partial/\partial y_j$: by employing the chain rule we have that

$$\omega_{ij}\partial_j K \;=\; \omega_{ij}\tilde{\partial}_l H\frac{\partial h_l}{\partial x_j} \tag{7.12}$$

(Here, as elsewhere in this book, a repeated index implies summation.) By virtue of (7.1) $\tilde{\partial}_l H = -\omega_{lm}\dot{y}_m$, so that, again by employing the chain rule, we obtain

$$\omega_{ij}\partial_j K \;=\; -\omega_{ij}\frac{\partial h_l}{\partial x_j}\omega_{lm}\frac{\partial h_m}{\partial x_n}\dot{x}_n \tag{7.13}$$

The right hand side simplifies to $\dot{x}_i$ (yielding Hamiltonian structure) only if

$$-\,\omega_{ij}\frac{\partial h_l}{\partial x_j}\omega_{lm}\frac{\partial h_m}{\partial x_n} \;=\; \delta_{in} \tag{7.14}$$

or, in compact notation, by defining $(\partial h)_{ij} = \frac{\partial h_i}{\partial x_j}$

$$-\,\omega(\partial h)^T\omega(\partial h) \;=\; \mathbf{1} \tag{7.15}$$

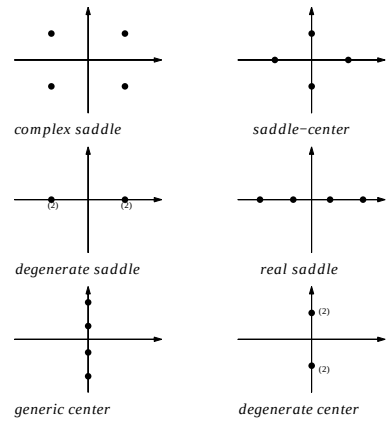



**Figure 7.3:** Stability exponents of a Hamiltonian equilibrium point, 2-dof.

which is equivalent to the requirement that $\partial h$ is symplectic. 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. [example 6.1]

**Example 7.3** ***Hamiltonian flows are canonical:*** *For Hamiltonian flows it follows from (7.10) that* $\frac{d}{dt}\left(J^T\omega J\right) = 0$, *and since at the initial time* $J^0(x_0) = \mathbf{1}$, *fundamental matrixis a symplectic transformation (7.11). 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 (7.11): For notational brevity here we have suppressed the dependence on time and the initial point,* $J = J^t(x_0)$. *By elementary properties of determinants it follows from (7.11) that Hamiltonian flows are phase space volume preserving:*

$$|\det J| = 1\,. \tag{7.16}$$

*Actually it turns out that for symplectic matrices (on any field) one always has* $\det J = +1$.

### 7.2.2 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 (7.10) with real matrix elements, so its eigenvalues (the Floquet exponents of (4.30)) are either real or come in complex pairs. In the case of Hamiltonian flows, it follows from (7.10) that the characteristic polynomial of $A$ for an equilibrium $x_q$ satisfies

$$\begin{aligned}\det(A-\lambda\mathbf{1}) &= \det(\omega^{-1}(A-\lambda\mathbf{1})\omega) = \det(-\omega A\omega-\lambda\mathbf{1})\\ &= \det(A^T+\lambda\mathbf{1}) = \det(A+\lambda\mathbf{1})\,.\end{aligned} \tag{7.17}$$

That is, the symplectic invariance implies in addition that if $\lambda$ is an eigenvalue, then $-\lambda$, $\lambda^*$ and $-\lambda^*$ are also eigenvalues. Distinct symmetry classes of the Floquet

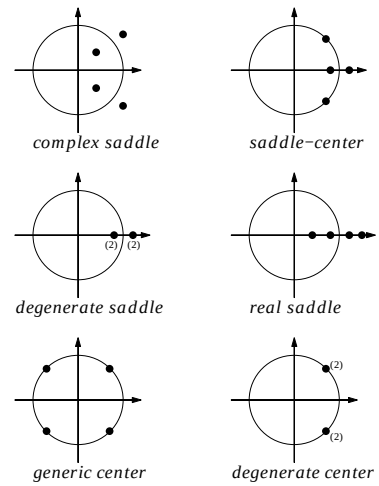



**Figure 7.4:** Stability of a symplectic map in $\mathbb{R}^4$.

exponents of an equilibrium point in a 2-dof system are displayed in figure 7.3. 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.

[section 4.3.1]
[exercise 7.4]
[exercise 7.5]

## 7.3 Symplectic maps

A stability eigenvalue $\Lambda = \Lambda(x_0, t)$ associated to a trajectory is an eigenvalue of the fundamental matrix $J$. As $J$ is symplectic, (7.11) implies that

$$J^{-1} = -\omega J^T \omega \,, \tag{7.18}$$

so the characteristic polynomial is reflexive, namely it satisfies

$$\begin{aligned} \det(J - \Lambda \mathbf{1}) &= \det(J^T - \Lambda \mathbf{1}) = \det(-\omega J^T \omega - \Lambda \mathbf{1}) \\ &= \det(J^{-1} - \Lambda \mathbf{1}) = \det(J^{-1}) \det(\mathbf{1} - \Lambda J) \\ &= \Lambda^{2D} \det(J - \Lambda^{-1} \mathbf{1}) \,. \end{aligned} \tag{7.19}$$

Hence if $\Lambda$ is an eigenvalue of $J$, so are $1/\Lambda$, $\Lambda^*$ and $1/\Lambda^*$. Real eigenvalues always come paired as $\Lambda$, $1/\Lambda$. The Liouville conservation of phase space volumes (7.16) is an immediate consequence of this pairing up of eigenvalues. The complex eigenvalues come in pairs $\Lambda$, $\Lambda^*$, $|\Lambda| = 1$, or in loxodromic quartets $\Lambda$, $1/\Lambda$, $\Lambda^*$ and $1/\Lambda^*$. These possibilities are illustrated in figure 7.4.

**Example 7.4** ***Hamiltonian Hénon map, reversibility:*** *By (4.53) the Hénon map (3.18) for $b = -1$ value is the simplest 2-d orientation preserving area-preserving map, often studied to better understand topology and symmetries of Poincaré sections of 2 dof 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*

$$x_{i+1} + x_{i-1} = a - x_i^2 \,, \quad i = 1, ..., n_p \,, \tag{7.20}$$

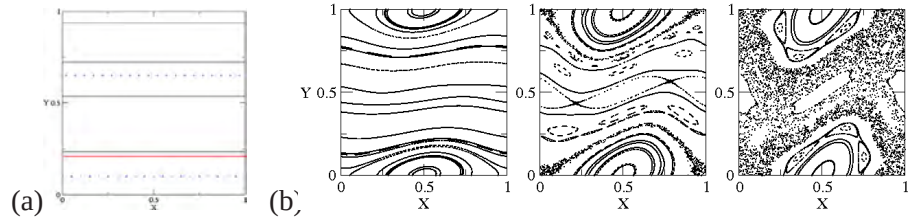



**Figure 7.5:** Phase portrait for the standard map for (a) $k = 0$: symbols denote periodic orbits, full lines represent quasiperiodic orbits. (b) $k = 0.3$, $k = 0.85$ and $k = 1.4$: each plot consists of 20 random initial conditions, each iterated 400 times.

*The 2-dimensional* Hénon map *for $b = -1$ parameter value*

$$\begin{aligned} x_{n+1} &= a - x_n^2 - y_n \\ y_{n+1} &= x_n \,. \end{aligned} \tag{7.21}$$

*is Hamiltonian (symplectic) in the sense that it preserves area in the $[x, y]$ plane.*

*For definitiveness, in numerical calculations in examples to follow 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.*

[exercise 8.6]

**Example 7.5** ***2-dimensional symplectic maps:*** *In the 2-dimensional case the eigenvalues (5.5) depend only on* $\mathrm{tr}\, M^t$

$$\Lambda_{1,2} = \frac{1}{2}\left(\mathrm{tr}\, M^t \pm \sqrt{(\mathrm{tr}\, M^t - 2)(\mathrm{tr}\, M^t + 2)}\right) . \tag{7.22}$$

*The trajectory is elliptic if the* stability residue $|\mathrm{tr}\, M^t| - 2 \leq 0$, *with complex eigenvalues* $\Lambda_1 = e^{i\theta t}$, $\Lambda_2 = \Lambda_1^* = e^{-i\theta t}$. *If* $|\mathrm{tr}\, M^t| - 2 > 0$, $\lambda$ *is real, and the trajectory is either*

$$\text{hyperbolic} \qquad \Lambda_1 = e^{\lambda t}\,, \quad \Lambda_2 = e^{-\lambda t}\,, \text{ or} \tag{7.23}$$

$$\text{inverse hyperbolic} \qquad \Lambda_1 = -e^{\lambda t}\,, \quad \Lambda_2 = -e^{-\lambda t}\,. \tag{7.24}$$

**Example 7.6** ***Standard map.*** *Given a smooth function $g(x)$, the map*

$$\begin{aligned} x_{n+1} &= x_n + y_{n+1} \\ y_{n+1} &= y_n + g(x_n) \end{aligned} \tag{7.25}$$

*is an area-preserving map. The corresponding $n$th iterate fundamental matrix (4.48) is*

$$M^n(x_0, y_0) = \prod_{k=n}^{1} \begin{pmatrix} 1 + g'(x_k) & 1 \\ g'(x_k) & 1 \end{pmatrix} . \tag{7.26}$$

*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 fancy way to say "circle"): by taking (7.25) mod $1$ the map can be reduced on the 2-torus $S_2$.*

*The* standard map *corresponds to the choice $g(x) = k/2\pi \sin(2\pi x)$. When $k = 0$, $y_{n+1} = y_n = y_0$, so that angular momentum is conserved, and the angle $x$ rotates with uniform velocity*

$$x_{n+1} = x_n + y_0 = x_0 + (n+1)y_0 \qquad \textit{mod } 1\,. \tag{7.27}$$

*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 7.5 (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 7.5 (a), while, when $k$ is sufficiently large, single trajectories wander erratically on a large fraction of the phase space, as in figure 7.5 (b).*

*This gives a glimpse of the typical scenario of transition to chaos for Hamiltonian systems.*

*Note that the map (7.25) provides a stroboscopic view of the flow generated by a (time-dependent) Hamiltonian*

$$H(x, y; t) = \frac{1}{2}y^2 + G(x)\delta_1(t) \tag{7.28}$$

*where $\delta_1$ denotes the periodic delta function*

$$\delta_1(t) = \sum_{m=-\infty}^{\infty} \delta(t - m) \tag{7.29}$$

*and*

$$G'(x) = -g(x) . \tag{7.30}$$

*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:*

$$R_{P/Q} = \frac{1}{4}(2 - \operatorname{tr} M_{P/Q}) . \tag{7.31}$$

*If $R_{P/Q} \in (0, 1)$ the orbit is elliptic, for $R_{P/Q} > 1$ the orbit is hyperbolic orbits, and for $R_{P/Q} < 0$ inverse hyperbolic.*

*For $k = 0$ all points on the $y_0 = P/Q$ line are periodic with period $Q$, winding number $P/Q$ and marginal stability $R_{P/Q} = 0$. As soon as $k > 0$, only a $2Q$ of such orbits survive, according to Poincaré-Birkhoff theorem: half of them elliptic, and half 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.*

## 7.4 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 (7.16) we derive the *Liouville theorem*:

$$
\begin{aligned}
V(t) &= \int_{f^t(C)} dx = \int_C \left| \det \frac{\partial f^t(x')}{\partial x} \right| dx' \\
& \int_C \det(J) dx' = \int_C dx' = V(0)\,, \qquad (7.32)
\end{aligned}
$$

Hamiltonian flows preserve phase space volumes.

The symplectic structure of Hamilton's equations buys us much more than the 'incompressibility,' or the phase space volume conservation. Consider the symplectic product of two infinitesimal vectors

$$
\begin{aligned}
(\delta x, \delta \hat{x}) &= \delta x^T \omega \delta \hat{x} = \delta p_i \delta \hat{q}_i - \delta q_i \delta \hat{p}_i \\
&= \sum_{i=1}^{D} \{\text{oriented area in the } (q_i, p_i) \text{ plane}\}\,. \qquad (7.33)
\end{aligned}
$$

Time $t$ later we have

$$
(\delta x', \delta \hat{x}') = \delta x^T J^T \omega J \delta \hat{x} = \delta x^T \omega \delta \hat{x}\,.
$$

This has the following geometrical meaning. We imagine there is a reference phase space point. We then define two other points infinitesimally close so that the vectors $\delta x$ and $\delta \hat{x}$ describe their displacements relative to the reference point. Under the 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 parallelopiped 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 $D$ oriented areas $\mathcal{V}_i$ bounded by $D$ loops $\Omega\mathcal{V}_i$, one per each $(q_i, p_i)$ plane, are separately conserved:

$$
\int_{\mathcal{V}} dp \wedge dq = \oint_{\Omega\mathcal{V}} p \cdot dq = \text{invariant}\,. \qquad (7.34)
$$

Morally a Hamiltonian flow is really $D$-dimensional, even though its phase space is $2D$-dimensional. Hence for Hamiltonian flows one emphasizes $D$, the number of the degrees of freedom.

in depth:
appendix B.3, p. 659

## Commentary

**Remark 7.1** Hamiltonian dynamics literature. 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? Out there there are about 2 centuries of accumulated literature on Hamilton, Lagrange, 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 [10] and Ozorio de Almeida [11].

**Remark 7.2** Symplectic. The term symplectic –Greek for twining or plaiting 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.

**Remark 7.3** The sign convention of $\omega$. The overall sign of $\omega$, the symplectic invariant in (7.7), is set by the convention that the 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 classical path is minimal, and the kinetic energy of a free particle is positive.

**Remark 7.4** Symmetries of the symbol square. For a more detailed discussion of symmetry lines see refs. [5, 8, 46, 13]. It is an open question (see remark 19.3) as to how time reversal symmetry can be exploited for reductions of cycle expansions. For example, the fundamental domain symbolic dynamics for reflection symmetric systems is discussed in some detail in sect. 19.5, but how does one recode from time-reversal symmetric symbol sequences to desymmetrized 1/2 state space symbols?

**Remark 7.5** Standard map. Standard maps model free rotators 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 exhibit a number of important features: small $k$ values provide an example of *KAM* perturbative regime (see ref. [8]), while for larger $k$ chaotic deterministic transport is observed [9, 10]; the transition to global chaos also presents remarkable universality features [11, 12, 13]. Also the quantum counterpart of this model has been widely investigated, being the first example where phenomena like quantum dynamical localization have been observed [14]. For some hands-on experience of the standard map, download Meiss simulation code [4].

## Exercises

7.1. **Complex nonlinear Schrödinger equation.** Consider the complex nonlinear Schrödinger equation in one spatial dimension [1]:

$$i\frac{\partial\phi}{\partial t} + \frac{\partial^2\phi}{\partial x^2} + \beta\phi|\phi|^2 = 0, \quad \beta \neq 0.$$

(a) Show that the function $\psi : \mathbb{R} \to \mathbf{C}$ 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)

7.2. **Symplectic group/algebra**

Show that if a matrix $C$ satisfies (7.10), then $\exp(sC)$ is a symplectic matrix.

7.3. **When is a linear transformation canonical?**

(a) Let $A$ be a $[n \times n]$ invertible matrix. Show that the map $\phi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ given by $(\mathbf{q}, \mathbf{p}) \mapsto (A\mathbf{q}, (A^{-1})^T\mathbf{p})$ 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)

7.4. **Determinant of symplectic matrices.** Show that the determinant of a symplectic matrix is +1, by going through the following steps:

(a) use (7.19) 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 $Q(1) = 0$).

7.5. **Cherry's example.** What follows refs. [2, 3] 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

$$H = \frac{1}{2}(q_1^2 + p_1^2) - (q_2^2 + p_2^2) + \frac{1}{2}p_2(p_1^2 - q_1^2) - q_1 q_2 p_1.$$

(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 $\tau$:

$$q_1 = -\sqrt{2}\frac{\cos(t-\tau)}{t-\tau}, \quad q_2 = \frac{\cos 2(t-\tau)}{t-\tau},$$
$$p_1 = \sqrt{2}\frac{\sin(t-\tau)}{t-\tau}, \quad p_2 = \frac{\sin 2(t-\tau)}{t-\tau}.$$

These solutions clearly blow up in finite time; however they start at $t = 0$ at a distance $\sqrt{3}/\tau$ from the origin, so by choosing $\tau$ 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)

## References

[7.1] J.E. Marsden and T.S. Ratiu, *Introduction to Mechanics and Symmetry* (Springer, New York, 1994).

[7.2] T.M. Cherry, "Some examples of trajectories defined by differential equations of a generalized dynamical type," *Trans.Camb.Phil.Soc.* **XXIII**, 165 (1925).

[7.3] K.R. Meyer, "Counter-examples in dynamical systems via normal form theory," *SIAM Review* **28**, 41 (1986)

[7.4] J.D. Meiss, "Visual explorations of dynamics: the standard map," arXiv:0801.0883.

[7.5] D.G. Sterling, H.R. Dullin and J.D. Meiss, "Homoclinic bifurcations for the Hénon map," *Physica D* **134**, 153 (1999); arXiv:chao-dyn/9904019.

[7.6] H.R. Dullin, J.D. Meiss and D.G. Sterling, "Symbolic codes for rotational orbits," arXiv:nlin.CD/0408015.

[7.7] A. Gómez and J.D. Meiss, "Reversible polynomial automorphisms of the plane: the involutory case," *Phys. Lett. A* **312**, 49 (2003); arXiv:nlin.CD/0209055.

[7.8] J.V. José and E.J. Salatan, *Classical dynamics - A contemporary approach* (Cambridge University Press, Cambridge, 1998)

[7.9] B.V. Chirikov, "A universal instability of many-dimensional oscillator system," *Phys.Rep.* **52**, 265 (1979).

[7.10] J.D. Meiss, "Symplectic maps, variational principles, and transport," *Rev.Mod.Phys.* **64**, 795 (1992).

[7.11] J.M. Greene, "A method for determining a stochastic transition," *J. Math. Phys.* **20**, 1183 (1979).

[7.12] J.M. Greene, "Two-dimensional measure-preserving mappings," *J. Math. Phys.* **9**, 760 (1968)

[7.13] S.J. Shenker and L.P. Kadanoff, "Critical behavior of a KAM surface: I. Empirical results," *J.Stat.Phys.* **27**, 631 (1982).

[7.14] G. Casati and B.V. Chirikov, *Quantum chaos: between order and disorder*, (Cambridge University Press, Cambridge, 1995)

# Chapter 8

# 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 intuitively most accessible example of chaos. Here we define billiard dynamics more precisely, anticipating the applications to come.

## 8.1 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 figure 8.1), 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 figure 1.1). In what follows we shall most often restrict our attention to *planar billiards*.

A point particle of mass $m$ and momentum $p_n = mv_n$ 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 momentum component normal to the boundary,

$$p' = p - 2(p \cdot \hat{n})\hat{n}\,, \tag{8.1}$$

with $\hat{n}$ the unit vector normal to the boundary $\partial Q$ at the collision point. The angle of incidence equals the angle of reflection, as illustrated in figure 8.2. A billiard is

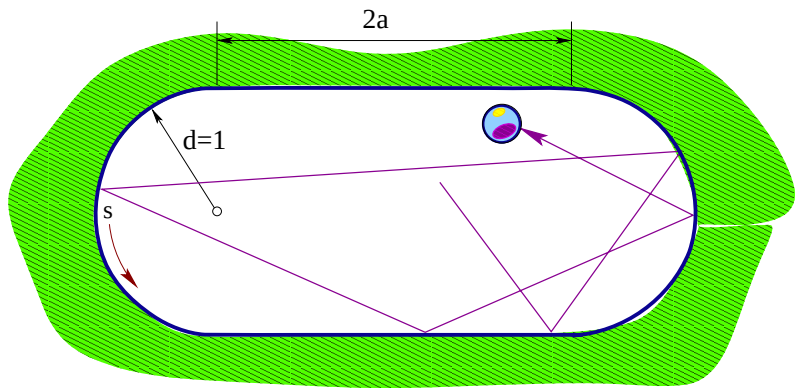



**Figure 8.1:** The stadium billiard is a 2-dimensional domain bounded by two semi-circles of radius $d = 1$ connected by two straight walls of length $2a$. At the points where the straight walls meet the semi-circles, the curvature of the border changes discontinuously; these are the only singular points of the flow. The length $a$ is the only parameter.

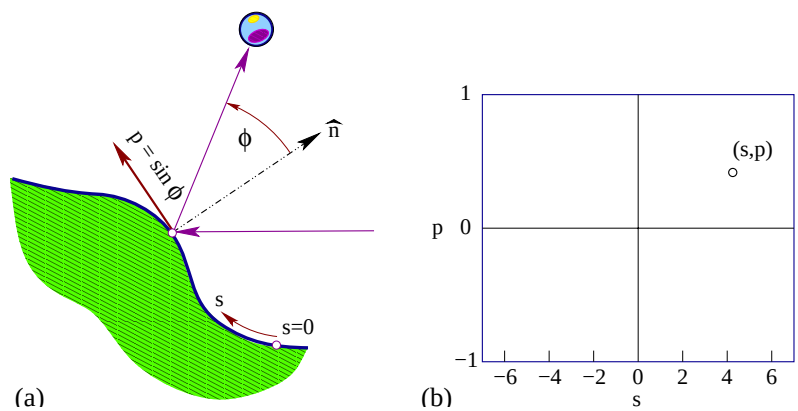



**Figure 8.2:** (a) A planar billiard trajectory is fixed by specifying the perimeter length parametrized by $s$ and the outgoing trajectory angle $\phi$, both measured counterclockwise with respect to the outward normal $\hat{n}$. (b) The Birkhoff phase space coordinate pair $(s, p)$ fully specifies 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 $q \in \partial Q$.

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 $\phi_n$ 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 8.2 as well as figure 3.3). In $D = 2$, the Poincaré section is a cylinder (topologically an annulus), figure8.3, where the parallel momentum $p$ ranges for $-|p|$ to $|p|$, and the $s$ coordinate is cyclic along each connected component of $\partial Q$. The volume in the full phase space is preserved by the Liouville theorem (7.32). The Birkhoff coordinates $x = (s, p) \in \mathcal{P}$, are the natural choice, because with them the the Poincaré return map preserves the phase space volume in the $(s, p)$ parameterize Poincaré section (a perfectly good coordinate set $(s, \phi)$ does not do that).

[exercise 8.6]
[section 8.2]

Without loss of generality we set $m = |v| = |p| = 1$. 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 Poincaré return map

$$P : (s_n, p_n) \mapsto (s_{n+1}, p_{n+1}) \tag{8.2}$$

from the $n$th collision to the $(n + 1)$st collision. The discrete time dynamics map

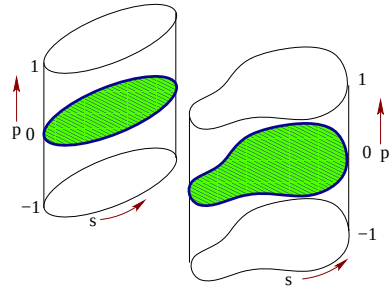



**Figure 8.3:** In $D = 2$ the billiard Poincaré section is a cylinder, with the parallel momentum $p$ ranging over $p \in \{-1, 1\}$, and with the $s$ coordinate is cyclic along each connected component of $\partial Q$. The rectangle figure 8.2 (b) is such cylinder unfolded, with periodic boundary conditions glueing together the left and the right edge of the rectangle.

$P$ is equivalent to the Hamiltonian flow (7.1) in the sense that both describe the same full trajectory. Let $t_n$ denote the instant of $n$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 $\tau$, the distance reached by the pinball along the $n$th section of its trajectory.

**Example 8.1** ***3-disk game of pinball:*** *In 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 3.3 we have two types of collisions:* [exercise 8.1]

$$P_0 : \begin{cases} \phi' = -\phi + 2\arcsin p \\ p' = -p + \frac{a}{R}\sin\phi' \end{cases} \qquad \textit{back-reflection} \tag{8.3}$$

$$P_1 : \begin{cases} \phi' = \phi - 2\arcsin p + 2\pi/3 \\ p' = p - \frac{a}{R}\sin\phi' \end{cases} \qquad \textit{reflect to 3rd disk}. \tag{8.4}$$

*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*'s either - one only needs to compute a square root per each reflection, and the simulations can be very fast.*

[exercise 8.2]

*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 on has to check whether the trajectory intersects the desired disk (and no disk inbetween). 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.*

## 8.2 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 (4.44) of the finite-time fundamental 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 the fact that the magnitude of the velocity is constant. We shall now show how going to a local frame of motion leads to a [2×2] fundamental matrix.

Consider a 2-dimensional billiard with phase space coordinates $x = (q_1, q_2, p_1, p_2)$. Let $t_k$ be the instant of the $k$th collision of the pinball with the billiard boundary, and $t_k^{\pm} = t_k \pm \epsilon$, $\epsilon$ positive and infinitesimal. With the mass and the velocity equal to 1, the momentum direction can be specified by angle $\theta$: $x = (q_1, q_2, \sin\theta, \cos\theta)$. Now parametrize the 2-$d$ neighborhood of a trajectory segment by $\delta x = (\delta z, \delta\theta)$, where

$$\delta z = \delta q_1 \cos\theta - \delta q_2 \sin\theta\,, \tag{8.5}$$

$\delta\theta$ is the variation in the direction of the pinball motion. Due to energy conservation, there is no need to keep track of $\delta q_{\|}$, variation along the flow, as that remains constant. $(\delta q_1, \delta q_2)$ is the coordinate variation transverse to the $k$th segment of the flow. From the 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

$$\frac{d}{dt}\delta\theta = 0, \qquad \frac{d}{dt}\delta z = \delta\theta\,. \tag{8.6}$$

Let $\delta\theta_k = \delta\theta(t_k^+)$ and $\delta z_k = \delta z(t_k^+)$ be the local coordinates immediately after the $k$th collision, and $\delta\theta_k^- = \delta\theta(t_k^-)$, $\delta z_k^- = \delta z(t_k^-)$ immediately before. Integrating the free flight from $t_{k-1}^+$ to $t_k^-$ we obtain

$$\begin{aligned} \delta z_k^- &= \delta z_{k-1} + \tau_k \delta\theta_{k-1}\,, \qquad \tau_k = t_k - t_{k-1} \\ \delta\theta_k^- &= \delta\theta_{k-1}\,, \end{aligned} \tag{8.7}$$

and the fundamental matrix (4.43) for the $k$th free flight segment is

$$M_T(x_k) = \begin{pmatrix} 1 & \tau_k \\ 0 & 1 \end{pmatrix}. \tag{8.8}$$

At incidence angle $\phi_k$ (the angle between the outgoing particle and the outgoing normal to the billiard edge), the incoming transverse variation $\delta z_k^-$ projects onto an arc on the billiard boundary of length $\delta z_k^-/\cos\phi_k$. The corresponding incidence angle variation $\delta\phi_k = \delta z_k^-/\rho_k \cos\phi_k$, $\rho_k$ = local radius of curvature, increases the angular spread to

$$\begin{aligned} \delta z_k &= -\delta z_k^- \\ \delta\theta_k &= -\delta\theta_k^- - \frac{2}{\rho_k \cos\phi_k}\delta z_k^-\,, \end{aligned} \tag{8.9}$$

so the fundamental matrix associated with the reflection is

$$M_R(x_k) = -\begin{pmatrix} 1 & 0 \\ r_k & 1 \end{pmatrix}, \qquad r_k = \frac{2}{\rho_k \cos\phi_k}\,. \tag{8.10}$$

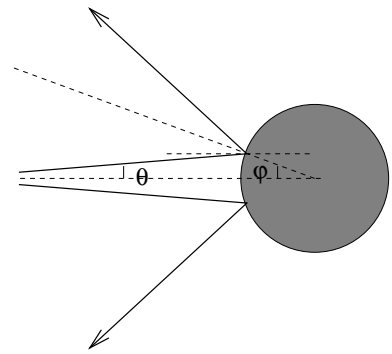



**Figure 8.4:** Defocusing of a beam of nearby trajectories at a billiard collision. (A. Wirzba)

The full fundamental matrix for $n_p$ consecutive bounces describes a beam of trajectories defocused by $M_T$ along the free flight (the $\tau_k$ terms below) and defocused/refocused at reflections by $M_R$ (the $r_k$ terms below)

[exercise 8.4]

$$M_p = (-1)^{n_p} \prod_{k=n_p}^{1} \begin{pmatrix} 1 & \tau_k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ r_k & 1 \end{pmatrix}, \tag{8.11}$$

where $\tau_k$ is the flight time of the $k$th free-flight segment of the orbit, $\eta_k = 2/\rho_k \cos\phi_k$ is the defocusing due to the $k$th reflection, and $\rho_k$ is the radius of curvature of the billiard boundary at the $k$th scattering point (for our 3-disk game of pinball, $\rho = 1$). As the billiard dynamics is phase space volume preserving, $\det M = 1$, and the eigenvalues are given by (7.22).

This is still another example of the fundamental matrix chain rule (4.51) for discrete time systems, rather similar to the Hénon map stability (4.52). Stability of every flight segment or reflection taken alone is a shear with two unit eigenvalues,

$$\det M_T = \det \begin{pmatrix} 1 & \tau_k \\ 0 & 1 \end{pmatrix}, \quad \det M_R = \det \begin{pmatrix} 1 & 0 \\ r_k & 1 \end{pmatrix}, \tag{8.12}$$

but acting in concert in the intervowen sequence (8.11) they can lead to a hyperbolic deformation of the infinitesimal neighborhood of a billiard trajectory.

[exercise 9.3]

As a concrete application, consider the 3-disk pinball system of sect. 1.3. Analytic expressions for the lengths and eigenvalues of $\overline{0}$, $\overline{1}$ and $\overline{10}$ cycles follow from elementary geometrical considerations. Longer cycles require numerical evaluation by methods such as those described in chapter 12.

[exercise 9.4]
[exercise 8.3]
[chapter 12]

## Résumé

A particulary natural application of the Poincaré section method is the reduction of a billiard flow to a boundary-to-boundary return map.

## Commentary

**Remark 8.1** Billiards. 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. 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 expansions curvature corrections that will be central to this treatise.

[chapter 18]

That the 3-disk game of pinball is a quintessential example of deterministic chaos appears to have been first noted by B. Eckhardt [1]. The model was studied in depth classically, semiclassically and quantum mechanically by P. Gaspard and S.A. Rice [3], 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 [5], construct semiclassical evolution operators and entire spectral determinants by P. Cvitanović and G. Vattay [6], and incorporate the diffraction effects into the periodic orbit theory by G. Vattay, A. Wirzba and P.E. Rosenqvist [7]. Gaspard's monograph [9], which we warmly recommend, utilizes the 3-disk system in much more depth than will be attained here. For further links check `ChaosBook.org`.

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 much harder challenge, 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 27.1]

**Remark 8.2** Stability analysis. The chapter 1 of Gaspard monograph [9] is recommended reading if you are interested in Hamiltonian flows, and billiards in particular. A. Wirzba has generalized the stability analysis of sect. 8.2 to scattering off 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 [9], sect. 1.4.

## Exercises

8.1. **A pinball simulator.** Implement the disk → 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 1.9 or figure 10.4 for $R$:$a$ = 2.5 and/or 6.

8.2. **Trapped orbits.** Shoot 100,000 trajectories from one of the disks, and trace out the strips of figure 1.9 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.

8.3. **Pinball stability.** Add to your exercise 8.1 pinball simulator a routine that computes the the [2×2] Jacobian matrix. To be able to compare with the numerical results of coming chapters, work with $R$:$a$ = 6 and/or 2.5 values.

8.4. **Stadium billiard.** Consider the *Bunimovich stadium* [9, 10] defined in figure 8.1. The fundamental matrix associated with the reflection is given by (8.10). 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 fundamental matrix of one semicircle reflection folowed by the flight to the next bounce is

$$\begin{aligned}\mathbf{J} &= (-1)\begin{pmatrix} 1 & 2\cos\phi_k \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -2/\cos\phi_k & 1 \end{pmatrix} \\ &= (-1)\begin{pmatrix} -3 & 2\cos\phi_k \\ 2/\cos\phi_k & 1 \end{pmatrix}.\end{aligned}$$

A shift must always be followed by $k = 1, 2, 3, \cdots$ bounces along a semicircle, hence the natural symbolic dynamics for this problem is $n$-ary, with the corresponding fundamental matrix given by shear (*ie.* the eigenvalues remain equal to 1 throughout the whole rotation), and $k$ bounces inside a circle lead to

$$\mathbf{J}^k = (-1)^k \begin{pmatrix} -2k-1 & 2k\cos\phi \\ 2k/\cos\phi & 2k-1 \end{pmatrix}. \tag{8.13}$$

The fundamental matrix of a cycle $p$ of length $n_p$ is given by

$$\mathbf{J}_p = (-1)^{\sum n_k} \prod_{k=1}^{n_p} \begin{pmatrix} 1 & \tau_k \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ n_k r_k & 1 \end{pmatrix}. \tag{8.14}$$

Adopt your pinball simulator to the stadium billiard.

8.5. **A test of your pinball simulator.** Test your exercise 8.3 pinball simulator by computing numerically cycle stabilities by tracking distances to nearby orbits. Compare your result with the exact analytic formulas of exercise 9.3 and 9.4.

8.6. **Birkhoff coordinates.** Prove that the Birkhoff coordinates are phase space volume preserving.

## References


[8.1] B. Eckhardt, *Fractal properties of scattering singularities*, *J. Phys.* **A 20**, 5971 (1987).

[8.2] G.D. Birkhoff, *Acta Math.* **50**, 359 (1927), reprinted in ref. [21].

[8.3] P. Gaspard and S.A. Rice, *J. Chem. Phys.* **90**, 2225 (1989); **90**, 2242 (1989); **90**, 2255 (1989).

[8.4] P. Cvitanović and B. Eckhardt, "Periodic-orbit quantization of chaotic system," *Phys. Rev. Lett.* **63**, 823 (1989).

[8.5] P. Gaspard and D. Alonso Ramirez, *Phys. Rev.* **A 45**, 8383 (1992).

[8.6] P. Cvitanović and G. Vattay, *Phys. Rev. Lett.* **71**, 4138 (1993).

[8.7] G. Vattay, A. Wirzba and P.E. Rosenqvist, *Phys. Rev. Lett.* **73**, 2304 (1994).

[8.8] Ya.G. Sinai, *Usp. Mat. Nauk* **25**, 141 (1970).

[8.9] L.A. Bunimovich, *Funct. Anal. Appl.* **8**, 254 (1974).

[8.10] L.A. Bunimovich, *Comm. Math. Phys.***65**, 295 (1979).

[8.11] L. Bunimovich and Ya.G. Sinai, *Markov Partition for Dispersed Billiard, Comm. Math. Phys.* **78**, 247 (1980); **78**, 479 (1980); *Erratum, ibid.* **107**, 357 (1986).

[8.12] R. Bridges, "The spin of a bouncing 'superball," *Phys. Educ.* **26**, 350 (1991); www.iop.org/EJ/abstract/0031-9120/26/6/003

[8.13] H. Lamba, "Chaotic, regular and unbounded behaviour in the elastic impact oscillator," chao-dyn/9310004

[8.14] S.W. Shaw and P.J. Holmes, *Phys. Rev. Lett.***51**, 623 (1983).

[8.15] C.R. de Oliveira and P.S. Goncalves, "Bifurcations and chaos for the quasiperiodic bouncing ball," *Phys. Rev. E* **56**, 4868 (1997).

[8.16] E. Cataldo and R. Sampaio, "A Brief Review and a New Treatment for Rigid Bodies Collision Models," *J. Braz. Soc. Mech. Sci.* **23** (2001).

[8.17] J. M. T. Thompson and R. Ghaffari. *Phys. Lett. A* **91**, 5 (1982).

[8.18] J.M.T. Thompson, A.R. Bokaian and R. Ghaffari. *J. Energy Resources Technology (Trans ASME)*, **106**, 191-198 (1984).

[8.19] E. Fermi. *Phys. Rev.* **75**, 1169 (1949).

[8.20] J. P. Cleveland, B. Anczykowski, i A. E. Schmid, and V. B. Elings. *Appl. Phys. Lett.* **72**, 2613 (1998).

[8.21] G. A. Tomlinson, *Philos. Mag* **7**, 905 (1929).

[8.22] T. Gyalog and H. Thomas, Z. *Phys. Lett. B* **104**, 669 (1997).

[8.23] J. Berg and G. A. D. Briggs. *Phys. Rev. B* **55**, 14899 (1997).

[8.24] J. Guckenheimer, P. J. Holmes. *J. Sound Vib.* **84**, 173 (1982).

[8.25] J. M. Luck, Anita Mehta *Phys. Rev. E* **48**, 3988 (1993).

[8.26] A. Valance, D. Bideau. *Phys. Rev. E* **57**, 1886 (1998).

[8.27] S.M. Hammel, J.A. Yorke, and C. Grebogi. *J. Complexity* **3**, 136 (1987).

[8.28] L. Maťyaś, R. Klages. *Physica D* **187**, 165 (2004).

# Chapter 9

# World in a mirror

A detour of a thousand pages starts with a single misstep.
—Chairman Miaw

DYNAMICAL SYSTEMS often come equipped with discrete symmetries, such as the reflection symmetries of various potentials. As we shall show here and in chapter 19, symmetries simplify the dynamics in a rather 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* $\mathcal{M}/G$. If the symmetry is continuous the dynamics is reduced to a lower-dimensional desymmetrized system $\mathcal{M}/G$, with "ignorable" coordinates eliminated (but not forgotten). In either case families of symmetry-related full state space cycles are replaced by fewer and often much shorter "relative" cycles. In presence of a symmetry the notion of a prime periodic orbit has to be reexamined: it is replaced by the notion of a *relative periodic orbit*, the shortest segment of the full state space cycle which tiles the cycle 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.

Familiarity with basic group-theoretic notions is assumed, with details relegated to appendix H.1. The erudite reader might prefer to skip the lengthy group-theoretic overture and go directly to $C_2 = D_1$ example 9.1 and example 9.2, and $C_{3v} = D_3$ example 9.3, backtrack as needed.

Our hymn to symmetry is a symphony in two movements: In this chapter 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 19.

## 9.1 Discrete symmetries

We show that a symmetry equates multiplets of equivalent orbits.

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.

**Definition:** A **finite group** consists of a set of elements

$$G = \{e, g_2, \ldots, g_{|G|}\} \tag{9.1}$$

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 $h \circ g = g \circ h = e$.

$|G|$, the number of elements, is called the *order* of the group.

**Definition: Coordinate transformations.** An *active* linear coordinate transformation $x \to \mathbf{T}x$ corresponds to a non-singular $[d \times d]$ matrix $\mathbf{T}$ that shifts the vector $x \in \mathcal{M}$ into another vector $\mathbf{T}x \in \mathcal{M}$. The corresponding *passive* coordinate transformation $f(x) \to \mathbf{T}^{-1} f(x)$ changes the coordinate system with respect to which the vector $f(x) \in \mathcal{M}$ is measured. Together, a passive and active coordinate transformations yield the map in the transformed coordinates:

$$\hat{f}(x) = \mathbf{T}^{-1} f(\mathbf{T}x)\,. \tag{9.2}$$

Linear action of a discrete group $G$ element $g$ on states $x \in \mathcal{M}$ is given by a finite non-singular $[d \times d]$ matrix $\mathbf{g}$, the linear *representation* of element $g \in G$. In what follows we shall indicate by bold face $\mathbf{g}$ the matrix representation of the action of group element $g \in G$ on the state space vectors $x \in \mathcal{M}$.

If the coordinate transformation $\mathbf{g}$ 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

$$g^m \equiv \underbrace{g \circ g \circ \ldots \circ g}_{m \text{ times}} = e \quad \to \quad |\det \mathbf{g}| = 1\,. \tag{9.3}$$

As the modulus of its determinant is unity, $\det \mathbf{g}$ is an $m$th root of 1.

A group is a *symmetry* of a dynamics if for every solution $f(x) \in \mathcal{M}$ and $g \in G$, $y = \mathbf{g} f(x)$ is also a solution:

**Definition: Symmetry of a dynamical system.** A dynamical system $(\mathcal{M}, f)$ is *invariant* (or *G-equivariant*) under a symmetry group $G$ if the "equations of motion" $f : \mathcal{M} \to \mathcal{M}$ (a discrete time map $f$, or the continuous flow $f^t$) from the $d$-dimensional manifold $\mathcal{M}$ into itself commute with all actions of $G$,

$$f(\mathbf{g}x) = \mathbf{g} f(x) \,. \tag{9.4}$$

Another way to state this is that the "law of motion" is invariant, i.e., retains its form in any symmetry-group related coordinate frame (9.2),

$$f(x) = \mathbf{g}^{-1} f(\mathbf{g}x) \,, \tag{9.5}$$

for *any* state $x \in \mathcal{M}$ and *any* finite non-singular $[d \times d]$ matrix representation $\mathbf{g}$ of element $g \in G$.

Why "equivariant"? A function $h(x)$ is said to be *G-invariant* if $h(x) = h(\mathbf{g}x)$ for all $g \in G$. The map $f : \mathcal{M} \to \mathcal{M}$ maps vector into a vector, hence a slightly different invariance condition $f(x) = \mathbf{g}^{-1} f(\mathbf{g}x)$. It is obvious from the context, but for verbal emphasis some like to distinguish the two cases by *in/equi*-variant. The key result of the representation theory of invariant functions is:

**Hilbert-Weyl theorem.** For a compact group $G$ there exist a finite $G$-invariant homogenous polynomial basis $\{u_1, u_2, \ldots, u_m\}$ such that any $G$-invariant polynomial can be written as a multinomial

$$h(x) = p(u_1(x), u_2(x), \ldots, u_m(x)) \,. \tag{9.6}$$

In practice, explicit construction of such basis does not seem easy, and we will not take this path except for a few simple low-dimensional cases. We prefer to apply the symmetry to the system as given, rather than undertake a series of nonlinear coordinate transformations that the theorem suggests.

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, $f^t(x) \cap \mathbf{g} f^{t'}(x) = \emptyset$ for all $t, t'$. For compact invariant sets, such as fixed points and periodic orbits, especially the short ones, the situation is very different.

### 9.1.1 Isotropy subgroups

The subset of points $\mathcal{M}_{x_0} \subset \mathcal{M}$ that belong to the infinite-time trajectory of a given point $x_0$ is called the *orbit* (or a *solution*) of $x_0$. An orbit is a *dynamically invariant* notion: it refers to the totality of states that can be reached from $x_0$, with the full state space $\mathcal{M}$ foliated into 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 strategy of this monograph is to populate the state space by a hierarchy of *compact invariant sets* (equilibria, periodic orbits, invariant tori, ...), each computable in a finite time. Orbits which are compact invariant sets we label by whatever alphabet we find convenient in a particular application: $EQ = x_{EQ} = \mathcal{M}_{EQ}$ for an equilibrium, $p = \mathcal{M}_p$ for a periodic orbit, etc..

The set of points $\mathbf{g}x$ generated by all actions $g \in G$ of the group $G$ is called the *group orbit* of $x \in \mathcal{M}$. If $G$ is a symmetry, intrinsic properties of an equilibrium (such as Floquet exponents) or a cycle $p$ (period, Floquet multipliers) and its image under a symmetry transformation $g \in G$ are equal. A symmetry thus reduces the number of dynamically distinct solutions $\mathcal{M}_{x_0}$ of the system. So we also need to determine the symmetry of a *solution*, as opposed to (9.5), the symmetry of the *system*.

**Definition: Isotropy subgroup.** Let $p = \mathcal{M}_p \subset \mathcal{M}$ be an orbit of the system. A set of group actions which maps an orbit into itself,

$$G_p = \{g \subseteq G : g\mathcal{M}_p = \mathcal{M}_p\}\,, \tag{9.7}$$

is called an *isotropy subgroup* of the solution $\mathcal{M}_p$. We shall denote by $G_p$ the maximal *isotropy* subgroup of $\mathcal{M}_p$. For a discrete subgroup

$$G_p = \{e, b_2, b_3, \ldots, b_h\} \subseteq G\,, \tag{9.8}$$

of order $h = |G_p|$, group elements (isotropies) map orbit points into orbit points reached at different times. For continuous symmetries the isotropy subgroup $G_p$ can be any continuous or discrete subgroup of $G$.

Let $H = \{e, b_2, b_3, \ldots, b_h\} \subseteq G$ be a subgroup of order $h = |H|$. The set of $h$ elements $\{c, cb_2, cb_3, \ldots, cb_h\}$, $c \in G$ but not in $H$, is called left *coset* $cH$. For a given subgroup $H$ the group elements are partitioned into $H$ and $m-1$ cosets, where $m = |G|/|H|$. The cosets cannot be subgroups, since they do not include the identity element.

### 9.1.2 Conjugate elements, classes and orbit multiplicity

If $G_p$ is the isotropy subgroup of orbit $\mathcal{M}_p$, elements of the coset space $g \in G/G_p$ generate the $m-1$ distinct copies of $\mathcal{M}_p$, so for discrete groups the multiplicity of an equilibrium or a cycle $p$ is $m_p = |G|/|G_p|$.

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 not-conjugate subsets called *classes*. The identity $e$ is always in the class $\{e\}$ of its own. This is the only class which is a subgroup, all other classes lack the identity element. Physical importance of classes is clear from (9.2), the way coordinate transformations act on mappings: action of elements of a class (say reflections, or discrete rotations) is equivalent up to a redefinition of the coordinate frame. We saw above that splitting of a group $G$ into an isotropy subgroup $G_p$ and $m-1$ cosets $cG_p$ relates a solution $\mathcal{M}_p$ to $m-1$ other distinct solutions $c\mathcal{M}_p$. Clearly all of them have equivalent isotropies: the precise statement is that the isotropy subgroup of orbit $c\,p$ is conjugate to the $p$ isotropy subgroup, $G_{c\,p} = c\,G_p\,c^{-1}$.

The next step is the key step; if a set of solutions is equivalent by symmetry (a circle, let's say), we would like to represent it by a single solution (shrink the circle to a point).

**Definition: Invariant subgroup.** A subgroup $H \subseteq G$ 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 $H$.

$H$ divides $G$ into $H$ and $m-1$ cosets, each of order $|H|$. Think of action of $H$ within each subset as identifying its $|H|$ elements as equivalent. This leads to the notion of $G/H$ as the *factor group* or *quotient group* $G/H$ of $G$, with respect to the *normal divisor* (or invariant subgroup) $H$. Its order is $m = |G|/|H|$, and its multiplication table can be worked out from the $G$ multiplication table class by class, with the subgroup $H$ playing the role of identity. $G/H$ is *homeomorphic* to $G$, with $|H|$ elements in a class of $G$ represented by a single element in $G/H$.

So far we have discussed the structure of a group as an abstract entity. Now we switch gears to what we really need this for: describe the action of the group on the state space of a dynamical system of interest.

**Definition: Fixed-point subspace.** The fixed-point subspace of a given subgroup $H \in G$, $G$ a symmetry of dynamics, is the set state space points left *pointwise* invariant under any subgroup action

$$\mathrm{Fix}(H) = \{x \in \mathcal{M} : \mathbf{h}\,x = x \text{ for all } h \in H\}\,. \tag{9.9}$$

A typical point in $\mathrm{Fix}(H)$ moves with time, but remains within $f(\mathrm{Fix}(H)) \subseteq \mathrm{Fix}(H)$ for all times. This suggests a systematic approach to seeking compact invariant solutions. The larger the symmetry subgroup, the smaller $\mathrm{Fix}(H)$, easing the numerical searches, so start with the largest subgroups $H$ first.

**Definition: Invariant subspace.** $\mathcal{M}_\alpha \subset \mathcal{M}$ is an *invariant* subspace if

$$\{\mathcal{M}_\alpha : \mathbf{g}x \in \mathcal{M}_\alpha \text{ for all } g \in G \text{ and } x \in \mathcal{M}_\alpha\}\,. \tag{9.10}$$

$\{0\}$ and $\mathcal{M}$ are always invariant subspaces. So is any Fix($H$) which is point-wise invariant under action of $G$. We can often decompose the state space into smaller invariant subspaces, with group acting within each "chunk" separately:

**Definition: Irreducible subspace.** A space $\mathcal{M}_\alpha$ whose only invariant subspaces are $\{0\}$ and $\mathcal{M}_\alpha$ is called *irreducible*.

As a first, coarse attempt at classification of orbits by their symmetries, we take note three types of equilibria or cycles: asymmetric $a$, symmetric equilibria or cycles $s$ built by repeats of relative cycles $\tilde{s}$, and boundary equilibria.

**Asymmetric cycles:** An equilibrium or periodic orbit is not symmetric if $\{x_a\} \cap \{\mathbf{g}x_a\} = \emptyset$, where $\{x_a\}$ is the set of periodic points belonging to the cycle $a$. Thus $g \in G$ generate $|G|$ distinct orbits with the same number of points and the same stability properties.

**Symmetric cycles:** A cycle $s$ is *symmetric* (or *self-dual*) if it has a non-trivial isotropy subgroup, i.e., operating with $g \in G_p \subset G$ on the set of cycle points reproduces the set. $g \in G_p$ acts a shift in time, mapping the cycle point $x \in \mathcal{M}_p$ into $f^{T_p/|G_p|}(x)$

**Boundary solutions:** An equilibrium $x_q$ or a larger compact invariant solution in a fixed-point subspace Fix($G$), $\mathbf{g}x_q = x_q$ for all $g \in G$ lies on the boundary of domains related by action of the symmetry group. A solution that is point-wise invariant under all group operations has multiplicity 1.

A string of unmotivated definitions (or an unmotivated definition of strings) has a way of making trite mysterious, so let's switch gears: develop a feeling for why they are needed by first working out the simplest, 1-$d$ example with a single reflection symmetry.

**Example 9.1** ***Group*** $D_1$ ***- a reflection symmetric*** $1d$ ***map:*** *Consider a 1d map $f$ with reflection symmetry $f(-x) = -f(x)$. An example is the bimodal "sawtooth" map of figure 9.1, piecewise-linear on 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. Denote the reflection operation by $Rx = -x$. The 2-element group $\{e, R\}$ goes by many names - here we shall refer to it as $C_2$, the group of rotations in the plane by angle $\pi$, or $D_1$, dihedral group with a single reflection. The symmetry invariance of the map implies that if $\{x_n\}$ is a trajectory, than also $\{Rx_n\}$ is a trajectory because $Rx_{n+1} = Rf(x_n) = f(Rx_n)$.*

**Asymmetric cycles:** *$R$ generates a reflection of the orbit with the same number of points and the same stability properties, see figure 9.1 (c).*

**Symmetric cycles:** *A cycle $s$ is* symmetric *(or* self-dual*) if operating with $R$ on the set of cycle points reproduces the set. The period of a symmetric cycle is even ($n_s = 2n_{\tilde{s}}$), and the mirror image of the $x_s$ cycle point is reached by traversing the irreducible segment $\tilde{s}$ (relative periodic orbit) of length $n_{\tilde{s}}$, $f^{n_{\tilde{s}}}(x_s) = Rx_s$, see figure 9.1 (b).*

**Boundary cycles:** *In the example at hand there is only one cycle which is neither symmetric nor antisymmetric, but lies on the boundary Fix($G$): the fixed point $\overline{C}$ at the origin.*

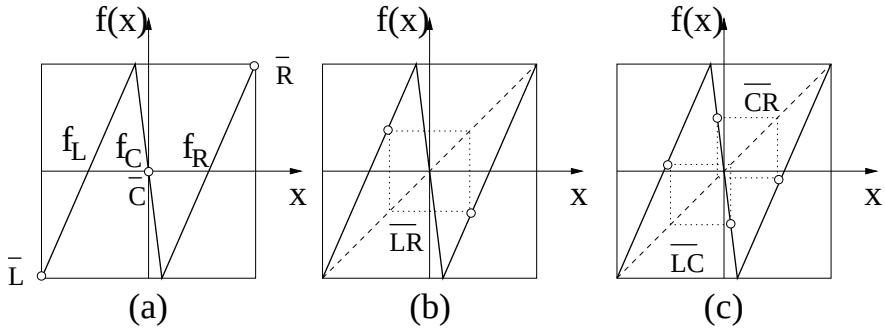



**Figure 9.1:** The bimodal Ulam sawtooth map with the $D_1$ symmetry $f(-x) = -f(x)$. (a) Boundary fixed point $\overline{C}$, asymmetric fixed points pair $\{\overline{L}, \overline{R}\}$. (b) Symmetric 2-cycle $\overline{LR}$. (c) Asymmetric 2-cycles pair $\{\overline{LC}, \overline{CR}\}$. Continued in figure 9.6). (Yueheng Lan)

*We shall continue analysis of this system in example 9.4, and work out the symbolic dynamics of such reflection symmetric systems in example 11.2.*

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-*d* maps.

For 3-*d* flows three types of discrete symmetry groups of order 2 can arise:

$$\begin{aligned} \text{reflections: } \sigma(x,y,z) &= (x,y,-z) \\ \text{rotations: } R(x,y,z) &= (-x,-y,z) \\ \text{inversions: } P(x,y,z) &= (-x,-y,-z) \end{aligned} \tag{9.11}$$

**Example 9.2** ***Desymmetrization of Lorenz flow:*** *(Continuation of example 4.7.) Lorenz equation (2.12) is invariant under the action of dihedral group $D_1 = \{e, R\}$, where $R$ is $[x,y]$-plane rotation by $\pi$ about the $z$-axis:*

$$(x,y,z) \to R(x,y,z) = (-x,-y,z)\,. \tag{9.12}$$

*$R^2 = 1$ 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*

$$\mathbf{P}^+ = \frac{1}{2}(1+R) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{P}^- = \frac{1}{2}(1-R) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{9.13}$$

*As the flow is $D_1$-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-d $\mathcal{M}^+$ subspace is the fixed-point subspace of $D_1$, with the $z$-axis points left* point-wise *invariant under the group action*

$$\mathit{Fix}(D_1) = \{x \in \mathcal{M}^+ : \boldsymbol{g}\,x = x \text{ for } g \in \{e, R\}\}\,. \tag{9.14}$$

*A point $x(t)$ in $\mathit{Fix}(G)$ moves with time, but remains within $x(t) \subseteq \mathit{Fix}(G)$ for all times; the subspace $\mathcal{M}^+ = \mathit{Fix}(G)$ is* flow invariant. *In case at hand this jargon is a bit of an overkill: clearly for $(x,y) = (0,0)$ the full state space Lorenz equation (2.12) is reduced to the exponential contraction to the $EQ_0$ equilibrium,*

$$\dot{z} = -b\,z\,. \tag{9.15}$$

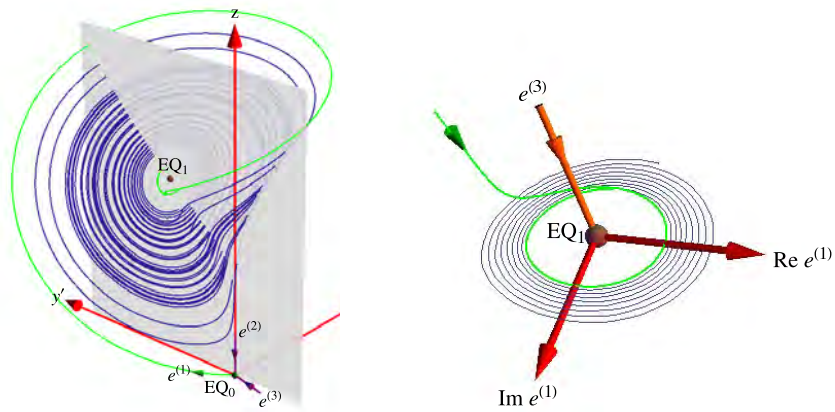



**Figure 9.2:** (a) Lorenz attractor plotted in $[x', y', z]$, the doubled-polar angle coordinates (9.16), with points related by $\pi$-rotation in the $[x, y]$ plane identified. Stable eigenvectors of $EQ_0$ $\mathbf{e}^{(3)}$ and $\mathbf{e}^{(2)}$, along the $z$ axis (9.15). Unstable manifold orbit $W^u(EQ_0)$ (green) is a continuation of the unstable $\mathbf{e}^{(1)}$ of $EQ_0$. (b) Blow-up of the region near $EQ_1$: The unstable eigenplane of $EQ_1$ is defined by Re $\mathbf{e}^{(2)}$ and Im $\mathbf{e}^{(2)}$, the stable eigenvector $\mathbf{e}^{(3)}$. 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. (E. Siminos)

*However, for flows in higher-dimensional state spaces the flow-invariant $\mathcal{M}^+$ subspace can itself be high-dimensional, with interesting dynamics of its own. Even in this simple case this subspace plays an important role as a topological obstruction, with the number of winds of a trajectory around it providing a natural symbolic dynamics.*

*The $\mathcal{M}^-$ subspace is, however,* not *flow-invariant, as the nonlinear terms $\dot{z} = xy - bz$ in the Lorenz equation (2.12) send all initial conditions within $\mathcal{M}^- = (x(0), y(0), 0)$ into the full, $z(t) \neq 0$ state space $\mathcal{M}$. The $R$ symmetry is nevertheless very useful.*

*By taking as a Poincaré section any $R$-invariant, infinite-extent, non-self-intersecting surface that contains the $z$ axis, the state space is divided into a half-space fundamental domain $\tilde{\mathcal{M}} = \mathcal{M}/D_1$ and its $180^o$ rotation $R\tilde{\mathcal{M}}$. An example is afforded by the $\mathcal{P}$ plane section of the Lorenz flow in figure 3.7. 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 $\pi$ around the $z$ axis.*

*As any such $R$-invariant section does the job, a choice of a "fundamental domain" is largely mater of taste. For purposes of visualization it is convenient to make instead 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:"

$$
\begin{aligned}
(x', y') &= (r\cos 2\theta, r\sin 2\theta) \\
&= ((x^2 - y^2)/r, 2xy/r)\,, \qquad (9.16)
\end{aligned}
$$

*as in figure 9.2 (a). In this representation the $\tilde{\mathcal{M}} = \mathcal{M}/D_1$ fundamental domain flow is a smooth, continuous flow, with (any choice of) the fundamental domain stretched out to seamlessly cover the entire $[x', y']$ plane.*

*We emphasize: such nonlinear coordinate transformations are* not *required to implement the symmetry quotienting $\mathcal{M}/G$, unless there are computational gains in a nonlinear coordinate change suggested by the symmetry. We offer them here only as a visualization aid that might help the reader disentangle 2-d projections of higher-dimensional flows. All numerical calculations are usually carried in the initial, full state space formulation of a flow, with symmetry-related points identified by* linear *symmetry transformations. (Continued in example 10.5.)*

*(E. Siminos and J. Halcrow)*

We now turn to discussion of a general discrete symmetry group, with elements that do not commute, and illustrate it by the 3-disk game of pinball, example 9.3 and example 9.5.

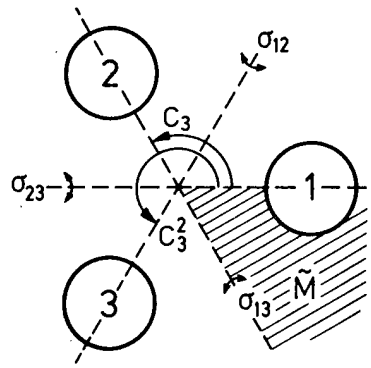



**Figure 9.3:** The symmetries of three disks on an equilateral triangle. The fundamental domain is indicated by the shaded wedge.

in depth:
appendix H, p. 699

## 9.2 Relative periodic orbits

We show that a symmetry reduces computation of periodic orbits to repeats of shorter, "relative periodic orbit" segments.

Invariance of a flow under a symmetry means that the symmetric 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 of the cycle) or it may be the same cycle.

A cycle is *symmetric* under symmetry operation $g$ if $g$ acts on it as a shift in time, advancing the starting point to the starting point of a symmetry related segment. A symmetric cycle $p$ can thus be subdivided into $m_p$ repeats of a *irreducible segment*, "prime" in the sense that the full state space cycle is a repeat of it. Thus in presence of a 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*.

Relative periodic orbits (or *equvariant periodic orbits*) are orbits $x(t)$ in state space $\mathcal{M}$ which exactly recur

$$x(t) = \mathbf{g}\, x(t+T) \tag{9.17}$$

for a fixed *relative period* $T$ and a fixed group action $g \in G$. This group action is referred to as a "phase," or a "shift." For a discrete group by (9.3) $g^m = e$ for some finite $m$, so the corresponding full state space orbit is periodic with period $mT$.

The period of the full orbit is given by the $m_p \times$(period of the relative periodic orbit), and the $i$th Floquet multiplier $\Lambda_{p,i}$ is given by $\Lambda_{\tilde{p},i}^{m_p}$ of the relative periodic orbit. The elements of the quotient space $b \in G/G_p$ generate the copies $bp$, so the multiplicity of the full state space cycle $p$ is $m_p = |G|/|G_p|$.

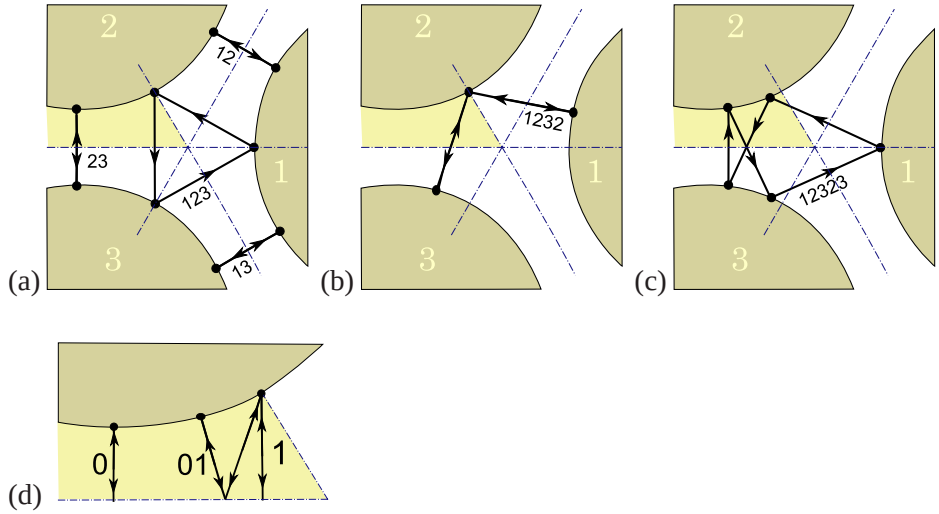



**Figure 9.4:** The 3-disk pinball cycles: (a) $\overline{12}$, $\overline{13}$, $\overline{23}$, $\overline{123}$. Cycle $\overline{132}$ turns clockwise. (b) Cycle $\overline{1232}$; the symmetry related $\overline{1213}$ and $\overline{1323}$ not drawn. (c) $\overline{12323}$; $\overline{12123}$, $\overline{12132}$, $\overline{12313}$, $\overline{13131}$ and $\overline{13232}$ not drawn. (d) The fundamental domain, i.e., the 1/6th wedge indicated 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 reduced to the two fixed points $\overline{0}$, $\overline{1}$, 2-cycle $\overline{10}$, and 5-cycle $\overline{00111}$ (not drawn).

We now illustrate these ideas with the example of sect. 1.3, symmetries of a 3-disk game of pinball.

**Example 9.3** $C_{3v} = D_3$ ***invariance - 3-disk game of pinball:*** *As the three disks in figure 9.3 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 $C_{3v}$ (dihedral group $D_3$), the group of order $|G| = 6$ consisting of the identity element $e$, three reflections across axes $\{\sigma_{12}, \sigma_{23}, \sigma_{13}\}$, and two rotations by $2\pi/3$ and $4\pi/3$ denoted $\{C, C^2\}$. 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, $\sigma_{12}$, the flip across the symmetry axis going through disk 1 interchanges the symbols 2 and 3; it maps the cycle $\overline{12123}$ into $\overline{13132}$, figure 9.5 (a). Cycles $\overline{12}$, $\overline{23}$, and $\overline{13}$ in figure 9.4 (a) are related to each other by rotation by $\pm 2\pi/3$, or, equivalently, by a relabeling of the disks.*

[exercise 9.6]

*The subgroups of $D_3$ are $D_1 = \{e, \sigma\}$, consisting of the identity and any one of the reflections, of order 2, and $C_3 = \{e, C, C^2\}$, of order 3, so possible cycle multiplicities are $|G|/|G_p| = 2, 3$ or 6.*

*The $C_3$ subgroup $G_p = \{e, C, C^2\}$ invariance is exemplified by 2 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 9.5 (b), and the multiplicity is $|G|/|G_p| = 2$.*

*The $C_v$ type of a subgroup is exemplified by the invariances of $\hat{p} = 1213$. This cycle is invariant under reflection $\sigma_{23}\{\overline{1213}\} = \overline{1312} = \overline{1213}$, so the invariant subgroup is $G_{\hat{p}} = \{e, \sigma_{23}\}$, with multiplicity is $m_{\hat{p}} = |G|/|G_p| = 3$; the cycles in this class, $\overline{1213}$, $\overline{1232}$ and $\overline{1323}$, are related by $2\pi/3$ rotations, figure 9.5 (c).*

*A cycle of no symmetry, such as $\overline{12123}$, 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 9.5 (a).*

*Besides the above discrete symmetries, for Hamiltonian systems cycles may be related by time reversal symmetry. An example are the cycles $\overline{121212313}$ and $\overline{121212323} = \overline{313212121}$ which have the same periods and stabilities, but are related by no space symmetry, see figure 9.5 (d). Continued in example 9.5.*

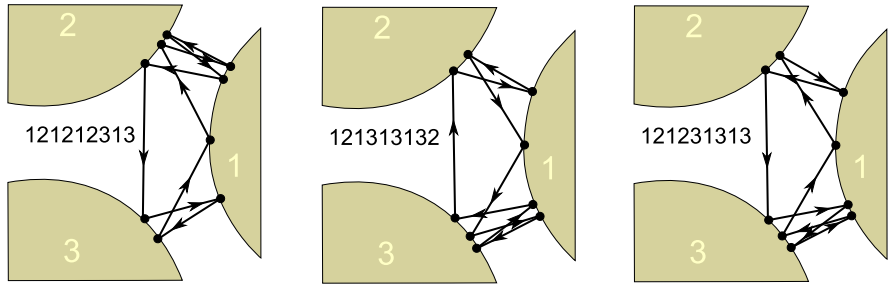



**Figure 9.5:** Cycle $\overline{121212313}$ has multiplicity 6; shown here is $\overline{121313132} = \sigma_{23}\overline{121212313}$. However, $\overline{121231313}$ which has the same stability and period is related to $\overline{121313132}$ by time reversal, but not by any $C_{3v}$ symmetry.

## 9.3 Domain for fundamentalists

So far we have used symmetry to effect a reduction in the number of independent cycles in cycle expansions. The next step achieves much more:

1. Discrete symmetries can be used to restrict all computations to a *fundamental domain*, the $\mathcal{M}/G$ quotiented subspace of the full state space $\mathcal{M}$.

2. Discrete symmetry tessellates the state space into copies of a fundamental domain, and thus induces a natural partition of state space. The state space is completely tiled by a fundamental domain and its symmetric images.

3. Cycle multiplicities induced by the symmetry are removed by *desymmetrization*, 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 $\tilde{p}$. 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 $h_{\tilde{p}}$. The relative periodic orbits in the full space, folded back into the fundamental domain, are periodic orbits.

4. The group elements $G = \{e, g_2, \ldots, g_{|G|}\}$ which map the fundamental domain $\tilde{\mathcal{M}}$ into its copies $g\tilde{\mathcal{M}}$, serve also as letters of a symbolic dynamics alphabet.

If the dynamics is invariant under a discrete symmetry, the state space $\mathcal{M}$ can be completely tiled by the fundamental domain $\tilde{\mathcal{M}}$ and its images $\mathcal{M}_a = a\tilde{\mathcal{M}}$, $\mathcal{M}_b = b\tilde{\mathcal{M}}, \ldots$ under the action of the symmetry group $G = \{e, a, b, \ldots\}$,

$$\mathcal{M} = \tilde{\mathcal{M}} \cup \mathcal{M}_a \cup \mathcal{M}_b \cdots \cup \mathcal{M}_{|G|} = \tilde{\mathcal{M}} \cup a\tilde{\mathcal{M}} \cup b\tilde{\mathcal{M}} \cdots . \tag{9.18}$$

Now we can use the invariance condition (9.4) to move the starting point $x$ into the fundamental domain $x = \mathbf{a}\tilde{x}$, and then use the relation $a^{-1}b = h^{-1}$ to also relate the endpoint $y$ to its image in the fundamental domain. 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.

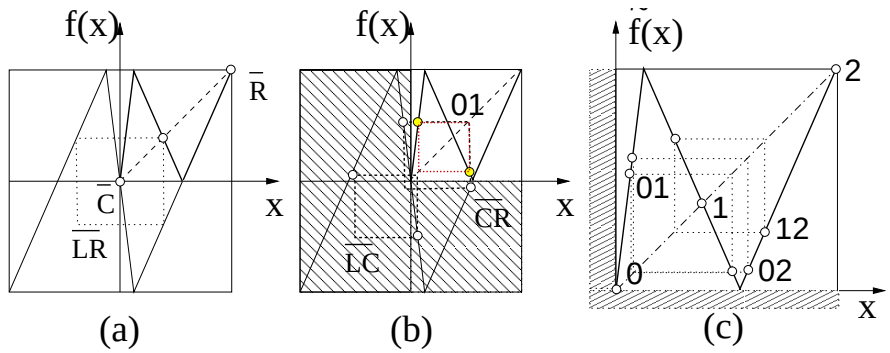



**Figure 9.6:** The bimodal Ulam sawtooth map of figure 9.1 with the $D_1$ symmetry $f(-x) = -f(x)$ restricted to the fundamental domain. $f(x)$ is indicated by the thin line, and fundamental domain map $\tilde{f}(\tilde{x})$ by the thick line. (a) Boundary fixed point $\overline{C}$ is the fixed point $\overline{0}$. The asymmetric fixed point pair $\{\overline{L},\overline{R}\}$ is reduced to the fixed point $\overline{2}$, and the full state space symmetric 2-cycle $\overline{LR}$ is reduced to the fixed point $\overline{2}$. (b) The asymmetric 2-cycle pair $\{\overline{LC},\overline{CR}\}$ is reduced to 2-cycle $\overline{02}$. (c) All fundamental domain fixed points and 2-cycles. (Yueheng Lan)

**Example 9.4** ***Group $D_1$ and reduction to the fundamental domain.*** *Consider again the reflection-symmetric bimodal Ulam sawtooth map $f(-x) = -f(x)$ of example 9.1, with symmetry group $D_1 = \{e, R\}$. The state space $\mathcal{M} = [-1, 1]$ can be tiled by half-line $\tilde{\mathcal{M}} = [0, 1]$, and $R\tilde{\mathcal{M}} = [-1, 0]$, 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 $R$.*

*In figure 9.6 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 orbits of the full state space dynamics are all periodic orbits in the fundamental domain. Consider figure 9.6*

*In (a) the boundary fixed point $\overline{C}$ is also the fixed point $\overline{0}$. In this case the set of points invariant under group action of $D_1$, $\tilde{\mathcal{M}} \cap R\tilde{\mathcal{M}}$, is just this fixed point $x = 0$, the reflection symmetry point.*

*The asymmetric fixed point pair $\{\overline{L},\overline{R}\}$ is reduced to the fixed point $\overline{2}$, and the full state space symmetric 2-cycle $\overline{LR}$ is reduced to the fixed point $\overline{1}$. The asymmetric 2-cycle pair $\{\overline{LC},\overline{CR}\}$ is reduced to the 2-cycle $\overline{01}$. Finally, the symmetric 4-cycle $\overline{LCRC}$ is reduced to the 2-cycle $\overline{02}$. This completes the conversion from the full state space for all fundamental domain fixed points and 2-cycles, figure 9.6 (c).*

**Example 9.5** ***3-disk game of pinball in the fundamental domain***

*If the dynamics is symmetric under interchanges of disks, the absolute disk labels $\epsilon_i = 1, 2, \cdots, N$ 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 $\sigma$ back to initial disk (symbol '0') or 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 9.4(d)). A set of orbits 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*

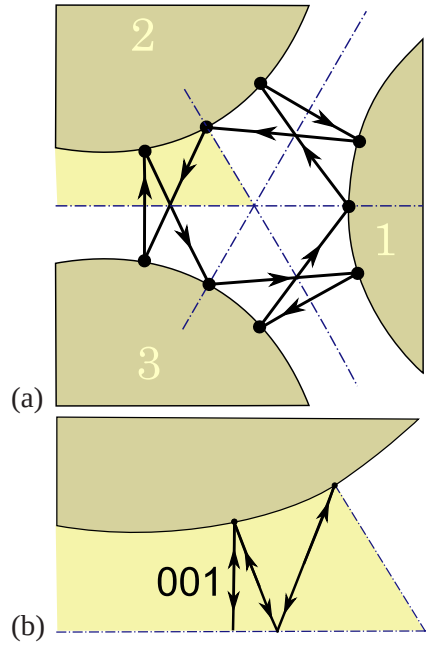



**Figure 9.7:** (a) The pair of full-space 9-cycles, the counter-clockwise $\overline{121232313}$ and the clockwise $\overline{131323212}$ correspond to (b) one fundamental domain 3-cycle $\overline{001}$.

*$\overline{12}$, $\overline{13}$ and $\overline{23}$) have multiplicity 3, and global orbits with no symmetry are 6-fold degenerate. Table **??** 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 9.5 and 9.7. Continued in example 11.3.*

## 9.4 Continuous symmetries

[...] which is an expression of consecration of "angular momentum."

— Mason A. Porter's student

What if the "law of motion" retains its form (9.5) in a family of coordinate frames $f(x) = \mathbf{g}^{-1} f(\mathbf{g}x)$ related by a group of *continuous* symmetries? The notion of "fundamental domain" is of no use here. Instead, as we shall see, continuous symmetries reduce dynamics to a desymmetrized system of lower dimensionality, by elimination of "ignorable" coordinates.

**Definition**: A Lie group is a topological group $G$ such that (1) $G$ has the structure of a smooth differential manifold. (2) The composition map $G \times G \to G : (g, h) \to gh^{-1}$ is smooth.

By "smooth" in this text we always mean $\mathbb{C}^\infty$ differentiable. If you are mystified by the above definition, don't be. Just think "aha, like the rotation group $SO(3)$?" If action of every element $g$ of a group $G$ commutes with the flow $\dot{x} = v(x)$, $x(t) = f^t(x_0)$,

$$\mathbf{g}v(x) = v(\mathbf{g}x)\,, \qquad \mathbf{g}f^t(x_0) = f^t(\mathbf{g}x_0)\,, \tag{9.19}$$

the dynamics is said to be *invariant* or *equivariant* under $G$.

Let $G$ be a group, $\mathcal{M}$ a set, and $g\mathcal{M} \longrightarrow \mathcal{M}$ a group action. For any $x \in \mathcal{M}$, the *orbit* $\mathcal{M}_x$ of $x$ is the set of all group actions

$$\mathcal{M}_x = \{\mathbf{g}\, x \mid g \in G\} \subset \mathcal{M}\,.$$

For a given state space point $x$ the group of $N$ continuous transformations together with the time translation sweeps out a smooth $(N+1)$-dimensional manifold of equivalent orbits. The time evolution itself is a noncompact 1-parameter Lie group; however, for solutions $p$ for which the $N$-dimensional group manifold is periodic in time $T_p$, the orbit of $x_p$ is a *compact* invariant manifold $\mathcal{M}_p$. The simplest example is the $N = 0$ case, where the invariant manifold $\mathcal{M}_p$ is the $1d$-torus traced out by the periodic trajectory. Thus the time evolution and the Lie group continuous symmetries can be considered on the same footing, and the closure of the set of compact unstable invariant manifolds $\mathcal{M}_p$ is the non–wandering set $\Omega$ of dynamics in presence of a continuous global symmetry (see sect.2.1.1).

The desymmetrized state space is the quotient space $\mathcal{M}/G$. The reduction to $\mathcal{M}/G$ amounts to a change of coordinates where the "ignorable angles" $\{t, \theta_1, \cdots, \theta_N\}$ parametrize $N+1$ time and group translations can be separated out. A simple example is the "rectification" of the harmonic oscillator by a change to polar coordinates, example 6.1.

### 9.4.1 Lie groups for pedestrians

All the group theory that you shall need here is in principle contained in the *Peter-Weyl theorem*, and its corollaries: A compact Lie group $G$ is completely reducible, its representations are fully reducible, 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.

Instead of writing yet another tome on group theory, in what follows we serve group theoretic nuggets on need-to-know basis, following a well-trod pedestrian route through a series of examples of familiar bits of group theory and Fourier analysis (but take a modicum of high, cyclist road in the text proper).

Consider infinitesimal transformations of form $g = 1 + iD$, $|D^a_b| \ll 1$, i.e., the transformations connected to the identity (in general, we also need to combine this with effects of invariance under discrete coordinate transformations, already discussed above). *Unitary* transformations $\exp(i\theta_j T_j)$ are generated by sequences of infinitesimal transformations of form

$$g_a{}^b \simeq \delta^a_b + i\delta\theta_i (T_i)^b_a \qquad \theta \in \mathbb{R}^N\,, \quad T_i \text{ hermitian}\,.$$

where $T_i$, the *generators* of infinitesimal transformations, are a set of linearly independent $[d \times d]$ hermitian matrices. In terms of the generators $T_i$, a tensor

$h_{ab\ldots}{}^{\ldots c}$ is invariant if $T_i$ "annihilate" it, i.e., $T_i \cdot h = 0$:

$$(T_i)_a^{a'} h_{a'b\ldots}{}^{c\ldots} + (T_i)_b^{b'} h_{ab'\ldots}{}^{c\ldots} - (T_i)_{c'}^{c} h_{ab\ldots}{}^{c'\ldots} + \ldots = 0\,. \tag{9.20}$$

**Example 9.6** ***Lie algebra.*** *As one does not want the symmetry rules to change at every step, the generators $T_i$, $i = 1, 2, \ldots, N$, are themselves invariant tensors:*

$$(T_i)\,{}_b^a = g^a{}_{a'} g_b{}^{b'} g_{ii'}\,(T_{i'})\,{}_{b'}^{a'}\,, \tag{9.21}$$

*where $g_{ij} = \left[e^{-i\theta_k C_k}\right]_{ij}$ is the adjoint $[N\times N]$ matrix representation of $g \in \mathcal{G}$. The $[d\times d]$ matrices $T_i$ are in general non-commuting, and from (9.20) it follows that they close $N$-element* Lie algebra

$$T_i T_j - T_j T_i = iC_{ijk} T_k \qquad i, j, k = 1, 2, ..., N\,,$$

*where the fully antisymmetric adjoint representation generators $[C_k]_{ij} = C_{ijk}$ are known as the* structure constants.

exercise 14.10

**Example 9.7** ***Group*** $SO(2)$**.** *$SO(2)$ is the group of rotations in a plane, smoothly connected to the unit element (i.e. the inversion $(x, y) \to (-x, -y)$ is excluded). A group element can be parameterized by angle $\theta$, and its action on smooth periodic functions is generated by*

$$g(\theta) = e^{i\theta \mathbf{T}}\,, \quad \mathbf{T} = -i\frac{d}{d\theta}\,,$$

*$g(\theta)$ rotates a periodic function $u(\theta + 2\pi) = u(\theta)$ by $\theta$ mod $2\pi$:*

$$g(\theta)u(\theta') = u(\theta' + \theta)$$

*The multiplication law is $g(\theta)g(\theta') = g(\theta + \theta')$. If the group $G$ actions consists of $N$ such rotations which commute, for example a $N$-dimensional box with periodic boundary conditions, the group $G$ is an Abelian group that acts on a torus $T^N$.*

### 9.4.2 Relative periodic orbits

Consider a flow invariant under a global continuous symmetry (Lie group) $G$. A relative periodic orbit $p$ is an orbit in state space $\mathcal{M}$ which exactly recurs

$$x_p(t) = \mathbf{g}_p x_p(t + T_p)\,, \qquad x_p(t) \in \mathcal{M}_p \tag{9.22}$$

for a fixed *relative period* $T_p$ and a fixed group action $g_p \in G$ that "rotates" the endpoint $x_p(T_p)$ back into the initial point $x_p(0)$. The group action $g_p$ is referred to as a "phase," or a "shift."

**Example 9.8** ***Continuous symmetries of the plane Couette flow.*** *The Navier-Stokes plane Couette flow defined as a flow between two countermoving planes, in a box periodic in streamwise and spanwise directions, a relative periodic solution is a solution that recurs at time $T_p$ with exactly the same disposition of velocity fields over the entire box, but shifted by a 2-dimensional (streamwise,spanwise) translation $g_p$. The $SO(2) \times SO(2)$ continuous symmetry acts on a 2-torus $T^2$.*

For dynamical systems with continuous symmetries parameters $\{t, \theta_1, \cdots, \theta_N\}$ are real numbers, ratios $\pi/\theta_j$ are almost never rational, and relative periodic orbits are almost never eventually periodic. As almost any such orbit explores ergodically the manifold swept by action of $G \times t$, they are sometimes referred to as "quasiperiodic." However, a relative periodic orbit can be pre-periodic if it is invariant under a discrete symmetry: 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, the orbit is periodic only after the action of $g$, as in (9.22).

In either discrete or continuous symmetry case, we refer to the orbits $\mathcal{M}_p$ in $\mathcal{M}$ satisfying (9.22) as *relative periodic orbits*. Morally, as it will be shown in chapter 19, they are the true "prime" orbits, i.e., the shortest segments that under action of $G$ tile the entire invariant submanifolds $\mathcal{M}_p$.

## 9.5 Stability

A infinitesimal symmetry group transformation maps a trajectory in a nearby equivalent trajectory, so we expect the initial point perturbations along to group manifold to be marginal, with unit eigenvalue. The argument is akin to (4.7), the proof of marginality of perturbations along a periodic orbit. In presence of an $N$-dimensional Lie symmetry group $G$, further $N$ eigenvalues equal unity. Consider two nearby initial points separated by an $N$-dimensional infinitesimal group transformation $\delta\theta$: $\delta x_0 = \mathbf{g}(\delta\theta)x_0 - x_0 = i\delta\theta \cdot \mathbf{T}x_0$. By the commutativity of the group with the flow, $\mathbf{g}(\delta\theta) f^t(x_0) = f^t(\mathbf{g}(\delta\theta)x_0)$. Expanding both sides, keeping the leading term in $\delta\theta$, and using the definition of the fundamental matrix (4.6), we observe that $J^t(x_0)$ transports the $N$-dimensional tangent vector frame at $x_0$ to the rotated tangent vector frame at $x(t)$ at time $t$:

$$\delta x(t) = \mathbf{g}(\theta) J^t(x_0)\, \delta x_0 \,. \tag{9.23}$$

For relative periodic orbits $\mathbf{g}_p x(T_p) = x(0)$, at any point along cycle $p$ the group tangent vector $\mathbf{T}x(t)$ is an eigenvector of the fundamental matrix $J_p(x) = \mathbf{g}_p J^{T_p}(x)$ with an eigenvalue of unit magnitude,

$$J^{T_p}(x)\, x_0 = \mathbf{g}(\theta)\mathbf{T}x(t)\,, \qquad x \in p\,. \tag{9.24}$$

Two successive points along the cycle separated by $\delta x_0$ have the same separation after a completed period $\delta x(T_p) = \mathbf{g}_p \delta x_0$, hence eigenvalue of magnitude 1.

### 9.5.1 Boundary orbits

Peculiar effects arise for orbits that run on a symmetry lines that border a fundamental domain. The state space transformation $\mathbf{h} \neq \mathbf{e}$ leaves invariant sets of *boundary* points; for example, under reflection $\sigma$ across a symmetry axis, the axis itself remains invariant. Some care need to be exercised in treating the invariant "boundary" set $\mathcal{M} = \tilde{\mathcal{M}} \cap \mathcal{M}_a \cap \mathcal{M}_b \cdots \cap \mathcal{M}_{|G|}$. The properties of boundary periodic orbits that belong to such pointwise invariant 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 in the bounded region of the Hénon-Heiles potential (remark9.3) and in $1d$ maps of example 9.1. 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 for low-dimensional state spaces there are typically relatively few boundary orbits, they tend to be among the shortest orbits, and they play a key role in dynamics.

While such boundary orbits are invariant under some symmetry operations, their neighborhoods are not. This affects the fundamental matrix $M_p$ of the orbit and its Floquet multipliers.

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 $\mathcal{M}/G$.

## Résumé

In sect. 2.1.1 we made a lame attempt to classify "all possible motions:" (1) equilibria, (2) periodic orbits, (3) everything else. Now one can discern in the fog of dynamics 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: (1) 0-dimensional equilibria $\mathcal{M}_q$, (2) 1-dimensional periodic orbits $\mathcal{M}_p$, (3) global symmetry induced $N$-dimensional relative equilibria $\mathcal{M}_{tw}$, (4) $(N+1)$-dimensional relative periodic orbits $\mathcal{M}_p$, (5) terra incognita. We have some inklings of the "terra incognita:" for example, symplectic symmetry induces existence of KAM-tori, and in general dynamical settings we are encountering more and more examples of *partially hyperbolic invariant tori*, isolated tori that are consequences of dynamics, not of a global symmetry, and which cannot be represented by a single relative periodic orbit, but require a numerical computation of full $(N+1)$-dimensional compact invariant sets and their infinite-dimensional linearized fundamental matrices, marginal in $(N+1)$ dimensions, and hyperbolic in the rest.

The main result of this chapter can be stated as follows: If a dynamical system $(\mathcal{M}, f)$ has a symmetry $G$, the symmetry should be deployed to "quotient" the state space $\mathcal{M}/G$, i.e., identify all $x \in \mathcal{M}$ related by the symmetry.

(1) In presence of a discrete symmetry $G$, associated with each full state space cycle $p$ is a maximal isotropy subgroup $G_p \subseteq G$ of order $1 \leq |G_p| \leq |G|$, whose elements leave $p$ invariant. The isotropy subgroup $G_p$ acts on $p$ as time shift, 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 $m_p = |G|/|G_p|$ distinct copies of $p$.

This reduction to the fundamental domain $\tilde{\mathcal{M}} = \mathcal{M}/G$ simplifies symbolic dynamics and eliminates symmetry-induced degeneracies. For the short orbits the labor saving is dramatic. For example, for the 3-disk game of pinball there are 256 periodic points of length 8, but reduction to the fundamental domain non-degenerate prime cycles reduces the number of the distinct cycles of length 8 to 30.

Amusingly, in this extension of "periodic orbit" theory from unstable 1-dimensional closed orbits to unstable $(N+1)$-dimensional compact manifolds $\mathcal{M}_p$ invariant under continuous symmetries, there are either no or proportionally few periodic orbits. Likelihood of finding a periodic orbit is *zero*. One expects some only if in addition to a continuous symmetry one has a discrete symmetry, or the particular invariant compact manifold $\mathcal{M}_p$ is invariant under a discrete subgroup of the continuous symmetry. 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 symmetries is a fool's errand.

Atypical as they are (no chaotic solution will be confined to these discrete subspaces) they are important for periodic orbit theory, as there the shortest orbits dominate.

We feel your pain, but trust us: once you grasp the relation between the full state space $\mathcal{M}$ and the desymmetrized $G$-quotiented $\mathcal{M}/G$, you will find the life as a fundamentalist so much simpler that you will never return to your full state space confused ways of yesteryear.

## Commentary

**Remark 9.1** Symmetries of the Lorenz equation: (Continued from remark 2.2.) After having studied example 9.2 you will appreciate why `ChaosBook.org` starts out with the symmetry-less Rössler flow (2.17), instead of the better known Lorenz flow (2.12) (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 manifolds of an equilibrium flowing into the stable manifold of another equilibria. How such connections are forced upon us is best grasped by perusing the chapter 13 "Heteroclinic tangles" of the inimitable Abraham and Shaw illustrated classic [26]. 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 [28] were first to quotient the $D_1$ 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 [33].

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 [29, 31]. They interpret the proto-Lorenz and its "double cover" Lorenz as "intensities" being the squares of "amplitudes," and call quotiented flows such as (Lorenz)/$D_1$ "images." Our "doubled-polar angle" visualization figure 10.7 is a proto-Lorenz in disguise, with the difference: we integrate the flow and construct Poincaré sections and return maps in the Lorenz $[x, y, z]$ coordinates, without any nonlinear coordinate transformations. The Poincaré return map figure 10.8 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 → itself [4], and thus intrinsic and coordinate independent. This is necessary in high-dimensional flows to avoid problems such as double-valuedness of return map projections on arbitrary 1–$d$ coordinates encountered already in the Rössler example. More importantly, as we know the embedding of the unstable manifold into the full state space, a cycle point of our return map *is* - regardless of the length of the cycle - the cycle point in the full state space, so no additional Newton searches are needed.

**Remark 9.2** Examples of systems with discrete symmetries. One has a $D_1$ symmetry in the Lorenz system (remark 2.2), the Ising model, and in the 3–$d$ anisotropic Kepler potential [4, 18, 19], a $D_3 = C_{3v}$ symmetry in Hénon-Heiles type potentials [5, 6, 7, 3], a $D_4 = C_{4v}$ symmetry in quartic oscillators [4, 5], in the pure $x^2y^2$ potential [6, 7] and in hydrogen in a magnetic field [8], and a $D_2 = C_{2v} = V_4 = C_2 \times C_2$ symmetry in the stadium billiard [9]. A very nice application of desymmetrization is carried out in ref. [10].

**Remark 9.3** Hénon-Heiles potential. An example of a system with $D_3 = C_{3v}$ symmetry is provided by the motion of a particle in the Hénon-Heiles potential [5]

$$V(r, \theta) = \frac{1}{2}r^2 + \frac{1}{3}r^3 \sin(3\theta) \ .$$

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 [8] discussed in sect. 9.5.1.

**Remark 9.4** Cycles and symmetries. We conclude this section with 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 surface of section. While in principle any Poincaré surface of section will do, a natural choice in the present context are crossings of symmetry axes, see example 7.6.

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 7.6.

## Exercises

9.1. **3-disk fundamental domain symbolic dynamics.** Try to sketch $\overline{0}$, $\overline{1}$, $\overline{01}$, $\overline{001}$, $\overline{011}$, $\cdots$. in the fundamental domain, and interpret the symbols $\{0,1\}$ by relating them to topologically distinct types of collisions. Compare with table **??**. 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, 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.

9.2. **Reduction of 3-disk symbolic dynamics to binary.**

(a) Verify that the 3-disk cycles
$\{\overline{12}, \overline{13}, \overline{23}\}$, $\{\overline{123}, \overline{132}\}$, $\{\overline{1213}+2 \text{ perms.}\}$,
$\{\overline{121232313}+5 \text{ perms.}\}$, $\{\overline{121323}+2 \text{ perms.}\}$,
$\cdots$,
correspond to the fundamental domain cycles $\overline{0}$, $\overline{1}$, $\overline{01}$, $\overline{001}$, $\overline{011}$, $\cdots$ respectively.

(b) Check the reduction for short cycles in table **??** by drawing them both in the full 3-disk system and in the fundamental domain, as in figure 9.7.

(c) Optional: Can you see how the group elements listed in table **??** relate irreducible segments to the fundamental domain periodic orbits?

9.3. **Fundamental domain fixed points.** Use the formula (8.11) for billiard fundamental matrix to compute the periods $T_p$ and the expanding eigenvalues $\Lambda_p$ of the fundamental domain $\overline{0}$ (the 2-cycle of the complete 3-disk space) and $\overline{1}$ (the 3-cycle of the complete 3-disk space) fixed points:

| | $T_p$ | $\Lambda_p$ |
|---|---|---|
| $\overline{0}$: | $R-2$ | $R-1+R\sqrt{1-2/R}$ |
| $\overline{1}$: | $R-\sqrt{3}$ | $-\frac{2R}{\sqrt{3}}+1-\frac{2R}{\sqrt{3}}\sqrt{1-\sqrt{3}/R}$ |

(9.25)

We have set the disk radius to $a=1$.

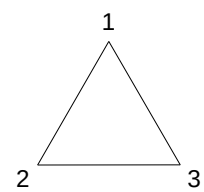
9.4. **Fundamental domain 2-cycle.** Verify that for the $\overline{10}$-cycle the cycle length and the trace of the fundamental matrix are given by

$$\begin{aligned} L_{10} &= 2\sqrt{R^2-\sqrt{3}R+1}-2, \\ \operatorname{tr}\mathbf{J}_{10} &= \Lambda_{10}+1/\Lambda_{10} \\ &= 2L_{10}+2+\frac{1}{2}\,\frac{L_{10}(L_{10}+2)^2}{\sqrt{3}R/2-1}\,. \end{aligned} \tag{9.26}$$

The $\overline{10}$-cycle is drawn in figure 11.2. The unstable eigenvalue $\Lambda_{10}$ follows from (7.22).

9.5. **A test of your pinball simulator: $\overline{10}$-cycle.** Test your exercise 8.3 pinball simulator stability evaluation by checking numerically the exact analytic $\overline{10}$-cycle stability formula (9.26).

9.6. **The group $C_{3v}$.** We will compute a few of the properties of the group $C_{3v}$, the group of symmetries of an equilateral triangle




(a) For this exercise, get yourself a good textbook, a book like Hamermesh [12] or Tinkham [11], and read up on classes and characters. All 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 $C_{3v}$ as cycles. For example, one of the rotations is (123), meaning that vertex 1 maps to 2 and 2 to 3 and 3 to 1.

(b) Find the subgroups of the group $C_{3v}$.

(c) Find the classes of $C_{3v}$ and the number of elements in them.

(d) There are three irreducible representations for the group. Two are one dimensional and the other one of multiplicity 2 is formed by $[2\times 2]$ matrices of the form

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Find the matrices for all six group elements.

(e) Use your representation to find the character table for the group.

9.7. **Lorenz system in polar coordinates: group theory.** Use (6.7), (6.8) to rewrite the Lorenz equation

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

in polar coordinates $(r,\theta,z)$, where $(x,y)=(r\cos\theta, r\sin\theta)$.

1. Show that in the polar coordinates Lorentz flow takes form

$$\begin{aligned}\dot{r} &= \frac{r}{2}\left(-\sigma - 1 + (\sigma+\rho-z)\sin 2\theta\right.\\ &\quad \left.+(1-\sigma)\cos 2\theta\right)\\ \dot{\theta} &= \frac{1}{2}\left(-\sigma+\rho-z+(\sigma-1)\sin 2\theta\right.\\ &\quad \left.+(\sigma+\rho-z)\cos 2\theta\right)\\ \dot{z} &= -bz + \frac{r^2}{2}\sin 2\theta\,. \end{aligned} \tag{9.27}$$

2. 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?

3. Show that this is the (Lorenz)$/D_1$ quotient map for the Lorenz flow, i.e., that it identifies points related by the $\pi$ rotation in the $(x, y)$ plane.

4. Show that a periodic orbit of the Lorenz flow in polar representation is either a periodic orbit or a relative periodic orbit (9.17) of the Lorenz flow in the $(x, y, z)$ representation.

5. Argue that if the dynamics is invariant under a rational rotation $R_{\pi/m}v(x) = v(R_{\pi/m}x) = v(x)$, a discrete subgroup $C_m$ of $SO(2)$ in the $(x, y)$-plane, the only non-zero Fourier components of equations of motion are $a_{jm} \neq 0$, $j = 1, 2, \cdots$. The Fourier representation is then the quotient map of the dynamics, $\mathcal{M}/C_m$.

By going to polar coordinates we have quotiented out the $\pi$-rotation $(x, y, z) \to (-x, -y, z)$ symmetry of the Lorenz equations, and constructed an explicit representation of the desymmetrized Lorenz flow.

9.8. **Lorenz system in polar coordinates: dynamics.** (Continuation of exercise 9.7.)

1. Show that (9.27) has two equilibria:

$$\begin{aligned}(r_0, z_0) &= (0, 0)\,, \qquad \theta_0 \text{ undefined}\\ (r_1, \theta_1, z_1) &= (\sqrt{2b(\rho-1)}, \pi/4, \rho-1)\end{aligned} \tag{9.28}$$

2. Verify numerically that the eigenvalues and eigenvectors of the two equilibria are:

   $EQ_1 = (0, 12, 27)$ equilibrium: (and its $R$-rotation related $EQ_2$ partner) 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
   $\mathrm{Re}\,\mathbf{e}^{(2)} = (-0.4955, -0.2010, -0.8450)$, $\mathrm{Im}\,\mathbf{e}^{(2)} = (0.5325, -0.8464, 0)$
   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 $\lambda^{(2)} = -b = -2.666\ldots$. The other stable eigenvector is
   $\mathbf{e}^{(2)} = (-0.244001, -0.969775, 0)$, with contracting eigenvalue $\lambda^{(2)} = -22.8277$. The unstable eigenvector
   $\mathbf{e}^{(3)} = (-0.653049, 0.757316, 0)$ has eigenvalue $\lambda^{(3)} = 11.8277$.

3. Plot the Lorenz strange attractor both in the original form figure 2.4 and in the doubled-polar coordinates (expand the angle $\theta \in [0, \pi]$ to $2\theta \in [0, 2\pi]$) 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 Poincaré return map $(s_n, s_n + 1)$ are plotted in figure 10.7.

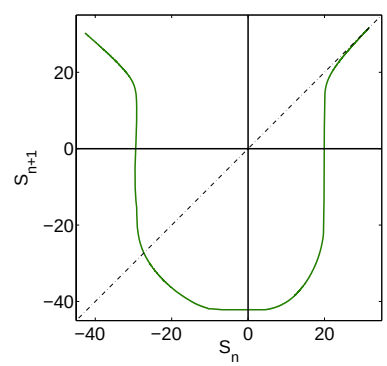



**Figure:** *The Poincaré return map $(s_n, s_{n+1})$ for the $EQ_0$, lower Poincaré section of figure 10.7 (b). (J. Halcrow)*

4. Construct the above Poincaré return map $(s_n, s_{n+1})$, where $s$ is arc-length measured along the unstable manifold of $EQ_0$. Elucidate its relation to the Poincaré return map of figure 10.8.

5. Show that if a periodic orbit of the polar representation Lorenz is also periodic orbit of the Lorenz flow, their stability eigenvalues are the same. How do the stability eigenvalues of relative periodic orbits of the representations relate to each other?

6. What does the volume contraction formula (4.34) look like now? Interpret.

9.9. **Proto-Lorenz system.** Here we quotient out the $D_1$ symmetry by constructing an explicit "intensity" representation of the desymmetrized Lorenz flow, following Miranda and Stone [28].

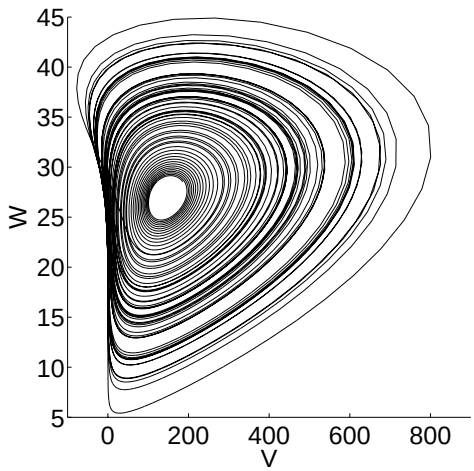



**Figure:** *The Lorenz attractor in proto-Lorenz representation (S.14). The points related by by $\pi$-rotation about the z-axis are identified. (J. Halcrow)*

1. Rewrite the Lorenz equation (2.12) in terms of variables

$$(u, v, z) = (x^2 - y^2, 2xy, z)\,, \tag{9.29}$$

show that it takes form

$$\begin{bmatrix} \dot{u} \\ \dot{v} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} -(\sigma+1)u + (\sigma - r)v + (1-\sigma)N + vz \\ (r-\sigma)u - (\sigma+1)v + (r+\sigma)N - uz - uN \\ v/2 - bz \end{bmatrix}$$

$$N = \sqrt{u^2 + v^2}\,.$$

2. Show that this is the (Lorenz)/$D_1$ quotient map for the Lorenz flow, i.e., that it identifies points related by the $\pi$ rotation (9.12).

3. Show that (9.29) 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 (2.12) and in the proto-Lorenz form for the Lorenz parameter values $\sigma = 10$, $b = 8/3$, $\rho = 28$, as in figure **??**. Topologically, does it resemble more the Lorenz, or the Rössler attractor, or neither?

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 stability eigenvalues are the same. How do the stability eigenvalues of relative periodic orbits of the Lorenz flow relate to the stability eigenvalues of the proto-Lorenz?

9 What does the volume contraction formula (4.34) look like now? Interpret.

10. Show that the coordinate change (9.29) is the same as rewriting (9.27) in variables $(u, v) = (r^2 \cos 2\theta, r^2 \sin 2\theta)$, i.e., squaring a complex number $z = x + iy$, $z^2 = u + iv$.

## References

[9.1] *Group theory - Birdtracks, Lie's, and Exceptional Groups*, www.birdtracks.eu (Princeton University Press 2008), in press

[9.2] P. Cvitanović and B. Eckhardt, "Symmetry decomposition of chaotic dynamics," *Nonlinearity* **6**, 277 (1993).

[9.3] G. Ott and G. Eilenberger, private communication.

[9.4] M.C. Gutzwiller, "The quantization of a classically ergodic system," *Physica* **D5**, 183 (1982).

[9.5] M. Henón and C. Heiles, *J. Astron.* **69**, 73 (1964).

[9.6] C. Jung and H.J. Scholz, *J. Phys.* **A 20**, 3607 (1987).

[9.7] C. Jung and P. Richter, *J. Phys.* **A 23**, 2847 (1990).

[9.8] B. Lauritzen, "Discrete symmetries and the periodic-orbit expansions," *Phys. Rev.* **A 43**, 603 (1991).

[9.9] J.M. Robbins, "Semiclassical trace formulas in the presence of continuous symmetries," *Phys. Rev.* **A 40**, 2128 (1989).

[9.10] N. Balasz and A. Voros, "Chaos on the pseudosphere," *Phys. Rep.* **143**, 109 (1986).

[9.11] M. Tinkham, *Group Theory and Quantum Mechanics* (McGraw-Hill, New York 1964).

[9.12] M. Hamermesh, *Group Theory and its Application to Physical Problems* (Addison-Wesley, Reading, 1962).

[9.13] E.N. Lorenz, "Deterministic nonperiodic flow," *J. Atmos. Sci.* **20**, 130 (1963).

[9.14] *Universality in Chaos*, P. Cvitanović, ed., (Adam Hilger, Bristol 1989).

[9.15] Bai-Lin Hao, *Chaos II* (World Scientific, Singapore, 1990).

[9.16] J. Frøyland, *Chaos and coherence* (Inst. of Phys. Publ., Bristol 1992).

[9.17] J. Frøyland and K.H. Alfsen, *Phys. Rev. A* **29**, 2928 (1984).

[9.18] Guckenheimer, J. and Williams, R., "Structural stability of the Lorenz attractor," *Publ. Math. IHES*, vol. 50, pp. 55–72, 1979.

[9.19] V. S. Afraimovich, B. B. Bykov, and L. P. Shilnikov, "On the appearence and the structure of the Lorenz attractor," *Dokl. Akad. Nauk SSSR* **234**, 336 (1987).

[9.20] B. Eckhardt and G. Ott, "Periodic orbit analysis of the Lorenz attractor," *J. Zeitschr. Physik B* **93**, 259 (1994).

[9.21] C. Sparrow, *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors* (Springer-Verlag, Berlin 1982).

[9.22] V. Franceschini, C. Giberti and Z.M. Zheng, "Characterization of the Lorenz attractor by unstable periodic orbits," *Nonlinearity* **6**, 251 (1993).

[9.23] B. Lahme and R. Miranda, "Karhunen-Loève Decomposition in the Presence of Symmetry - Part I," *IEEE TRANSACTIONS ON IMAGE PROCESSING* **8**, 1183 (1999)

[9.24] Jackson, E. A., *Perspectives of nonlinear dynamics: Vol. 1 and 2*. Cambridge: Cambridge University Press, 1989.

[9.25] Seydel, R., *From equilibrium to chaos: Practical bifurcation and stability analysis* (Elsevier, New York 1988).

[9.26] Abraham, R. H. and Shaw, C. D., *Dynamics - The geometry of behavior* (Addison-Wesley, Reading, MA 1992).

[9.27] M.L. Cartwright and J.E. Littlewood, "On nonlinear differential equations of the second order," *J. London Math. Soc.* **20**, 180 (1945).

[9.28] R. Miranda and E. Stone, "The proto-Lorenz system," '*Phys. Letters A* **178**, 105 (1993).

[9.29] R. Gilmore and C. Letellier, *The Symmetry of Chaos* (Oxford U. Press, Oxford 2007).

[9.30] C. Letellier, R. Gilmore and T. Jones, "Peeling Bifurcations of Toroidal Chaotic Attractors," arXiv:0707.3975v3.

[9.31] C. Letellier and R. Gilmore, "Covering dynamical systems: Two-fold covers," *Phys. Rev. E* **63**, 016206 (2001).

[9.32] R. Gilmore, "Two-parameter families of strange attractors," *Chaos* **17**, 013104 (2007).

[9.33] D. A. Cox, J. B. Little, and D. 0'Shea, *Ideals, Varieties and Algorithms* (Springer-Verlag, New York, 1996).

[9.34] F. Christiansen, P. Cvitanović and V. Putkaradze, "Hopf's last hope: spatiotemporal chaos in terms of unstable recurrent patterns," *Nonlinearity* **10**, 55 (1997),
chao-dyn/9606016.

[9.35] G. Tanner and D. Wintgen, "Quantization of chaotic systems." *CHAOS* **2**, 53 (1992).

[9.36] P. Cvitanović and F. Christiansen, "Periodic orbit quantization of the anisotropic Kepler problem," *CHAOS* **2**, 61 (1992).

# Chapter 10

# Qualitative dynamics, for pedestrians

> The classification of the constituents of a chaos, nothing less is here essayed.
>
> —Herman Melville, *Moby Dick*, chapter 32

In this chapter we begin to learn how to use qualitative properties of a flow in order to *partition* the state space in a topologically invariant way, and *name* topologically distinct orbits. This will enable us – in chapter 13 – 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.

We start by a simple physical example, symbolic dynamics of a 3-disk game of pinball, and then show that also for smooth flows the qualitative dynamics of stretching and folding flows enables us to partition the state space and assign symbolic dynamics itineraries to trajectories. Here we illustrate the method on a $1-d$ approximation to Rössler flow. In chapter 13 we turn this topological dynamics into a multiplicative operation on the state space partitions by means of transition matrices/Markov graphs, the simplest examples of evolution operators. Deceptively simple, this subject can get very difficult very quickly, so in this chapter we do the first pass, at a pedestrian level, postponing the discussion of higher-dimensional, cyclist level issues to chapter 11.

Even though by inclination you might only care about the serious stuff, like Rydberg atoms or mesoscopic devices, and resent wasting time on things formal, this chapter and chapter 13 are good for you. Read them.

## 10.1 Qualitative dynamics

(R. Mainieri and P. Cvitanović)

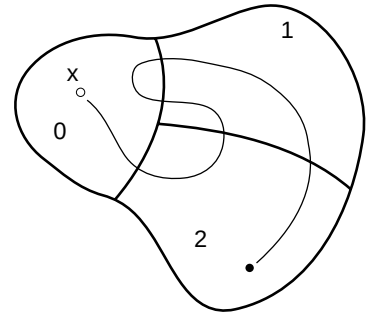



**Figure 10.1:** A trajectory with itinerary 021012.

What can a flow do to the state space points? This is a very difficult question to answer because we have assumed very little about the evolution function $f^t$; continuity, and differentiability a sufficient number of times. Trying to make sense of this question is one of the basic concerns in the study of dynamical systems. One of the first answers was inspired by the motion of the planets: they appear to repeat their motion through the firmament. Motivated by this observation, the first attempts to describe dynamical systems were to think of them 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

$$\left\{y \in \mathcal{M}_0 : \quad y = f^t(x_0), \quad t > t_0\right\} \tag{10.1}$$

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 turns out to be the important first step on the way to a theory of dynamical systems: qualitative, topological dynamics, or, as it is usually called, *symbolic dynamics*. As the subject can get quite technical, a summary of the basic notions and definitions of symbolic dynamics is relegated to sect. 10.5; check that section whenever you run into obscure symbolic dynamics jargon.

We start by cutting up the state space up into regions $\mathcal{M}_A, \mathcal{M}_B, \ldots, \mathcal{M}_Z$. This can be done in many ways, not all equally clever. Any such division of the state space into topologically distinct regions is a *partition*, and we associate with each region (sometimes referred to as a *state*) a symbol $s$ from an $N$-letter *alphabet* or *state set* $\mathcal{A} = \{A, B, C, \cdots, Z\}$. As the dynamics moves the point through the state space, different regions will be visited. The visitation sequence - forthwith referred to as the *itinerary* - can be represented by the letters of the alphabet $\mathcal{A}$. If, as in the example sketched in figure 10.1, the state space is divided into three regions $\mathcal{M}_0$, $\mathcal{M}_1$, and $\mathcal{M}_2$, the "letters" are the integers $\{0, 1, 2\}$, and the itinerary for the trajectory sketched in the figure is $0 \mapsto 2 \mapsto 1 \mapsto 0 \mapsto 1 \mapsto 2 \mapsto \cdots$.

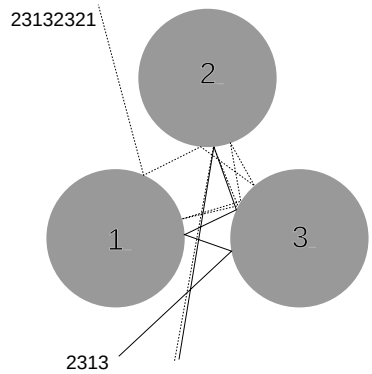



**Figure 10.2:** Two pinballs that start out very close to each other exhibit the same qualitative dynamics _2313_ for the first three bounces, but due to the exponentially growing separation of trajectories with time, follow different itineraries thereafter: one escapes after _2313_, the other one escapes after _23132321_.

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 partition is said to be *metrically indecomposable*.

In general one also encounters transient regions - regions to which the dynamics does not return to once they are exited. Hence we have to distinguish between (for us uninteresting) wandering trajectories that never return to the initial neighborhood, and the non–wandering set (2.2) of the *recurrent* trajectories.

The allowed transitions between the regions of a partition are encoded in the $[N\times N]$-dimensional *transition matrix* whose elements take values

$$T_{ij} = \begin{cases} 1 & \text{if a transition } \mathcal{M}_j \to \mathcal{M}_i \text{ is possible} \\ 0 & \text{otherwise .} \end{cases} \tag{10.2}$$

The transition matrix encodes the topological dynamics as an invariant law of motion, with the allowed transitions at any instant independent of the trajectory history, requiring no memory.

**Example 10.1** ***Complete $N$-ary dynamics:*** *All transition matrix entries equal unity (one can reach any region from any other region in one step):*

$$T_c = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} . \tag{10.3}$$

*Further examples of transition matrices, such as the 3-disk transition matrix (10.5) and the 1-step memory sparse matrix (10.13), are peppered throughout the text.*

However, knowing that a point from $\mathcal{M}_i$ reaches $\mathcal{M}_j$ in one step is not quite good enough. We would be happier if we knew that *any* point in $\mathcal{M}_i$ reaches $\mathcal{M}_j$; otherwise we have to subpartition $\mathcal{M}_i$ into the points which land in $\mathcal{M}_j$, and those which do not, and often we will find ourselves partitioning *ad infinitum*.

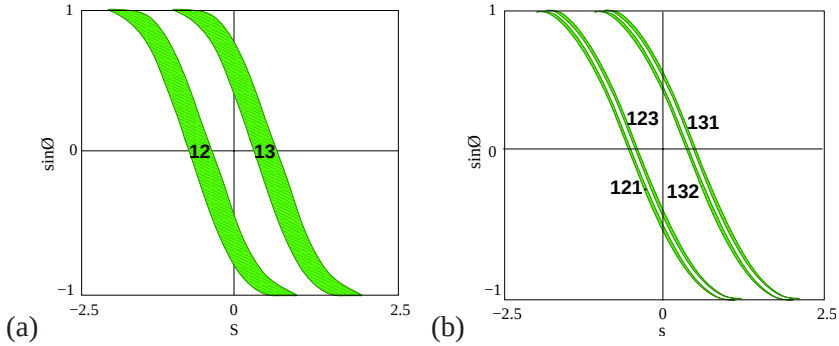



**Figure 10.3:** The 3-disk game of pinball Poincaré section, trajectories emanating from the disk 1 with $x_0$ = (arclength, parallel momentum) = $(s_0, p_0)$ , disk radius : center separation ratio a:R = 1:2.5. (a) Strips of initial points $\mathcal{M}_{12}$, $\mathcal{M}_{13}$ which reach disks 2, 3 in one bounce, respectively. (b) 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. (Y. Lan)

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. Dynamically, *finite Markov partitions* can be generated by *expanding* $d$-dimensional iterated mappings $f : \mathcal{M} \to \mathcal{M}$, if $\mathcal{M}$ can be divided into $N$ regions $\{\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_{N-1}\}$ such that in one step points from an initial region $\mathcal{M}_i$ either fully cover a region $\mathcal{M}_j$, or miss it altogether,

$$\text{either} \quad \mathcal{M}_j \cap f(\mathcal{M}_i) = \emptyset \quad \text{or} \quad \mathcal{M}_j \subset f(\mathcal{M}_i)\,. \tag{10.4}$$

Let us illustrate what this means by our favorite example, the game of pinball.

**Example 10.2** ***3-disk symbolic dynamics:*** *Consider the motion of a free point particle in a plane with* 3 *elastically reflecting convex disks. After a collision with a disk a particle either continues to another disk or escapes, and any trajectory can be labeled by the disk sequence. For example, if we label the three disks by* 1, 2 *and* 3, *the two trajectories in figure 10.2 have itineraries* _2313_, _23132321_ *respectively. The 3-disk prime cycles given in figures 9.4 and 11.2 are further examples of such itineraries.* [exercise 1.1]

*At each bounce a cone of initially nearby trajectories defocuses (see figure 1.8), and in order to attain a desired longer and longer itinerary of bounces the initial point* $x_0 = (s_0, p_0)$ *has to be specified with a larger and larger precision, and lie within initial state space strips drawn in figure 10.3. Similarly, it is intuitively clear that as we go backward in time (in this case, simply 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* $\sin\phi \to -\sin\phi$*).* $\mathcal{M}_{s_1.s_2}$, *figure 10.4, is a "rectangle" of nearby trajectories which have arrived from the disk* $s_1$ *and are heading for the disk* $s_2$.

The itinerary is finite for a scattering trajectory, coming in from infinity and escaping after a finite number of collisions, infinite for a trapped trajectory, and 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. More generally, for hyperbolic flows the intersection of the future and past itineraries, the bi-infinite itinerary $S^-.S^+ = \cdots s_{-2}s_{-1}s_0.s_1s_2s_3\cdots$ specifies a unique trajectory. This is intuitively clear for our 3-disk game of pinball, and is stated more formally in the definition (10.4) of a Markov partition. The definition requires that the dynamics be expanding forward in time in order to ensure that the cone of trajectories with a given itinerary becomes sharper and sharper as the number of specified symbols is increased.

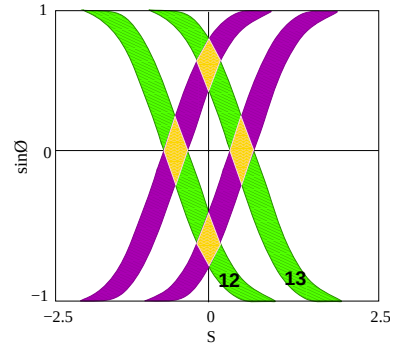



**Figure 10.4:** The Poincaré section of the state space for the binary labeled pinball. For definitiveness, this set is generated by starting from disk 1, preceded by disk 2. Indicated are the fixed points $\overline{0}$, $\overline{1}$ and the 2-cycle periodic points $\overline{01}$, $\overline{10}$, together with strips which survive 1, 2, ... bounces. Iteration corresponds to the decimal point shift; for example, all points in the rectangle [01.01] map into the rectangle [010.1] in one iteration. See also figure 11.2 (b).

**Example 10.3** ***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, *hence, the dynamics is a subshift of finite type (see (10.22) for definition), with the transition matrix (10.2) given by*

$$T = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}. \tag{10.5}$$

*For convex disks the separation between nearby trajectories increases at every reflection, implying that the fundamental matrix has an expanding eigenvalue. By the Liouville phase space volume conservation (7.32), the other transverse eigenvalue is contracting. This example demonstrates that finite Markov partitions can be constructed for* hyperbolic *dynamical systems which are expanding in some directions, contracting in others. Further examples are the* 1*-dimensional expanding mapping sketched in figure 10.6, and more examples are worked out in sect. 24.2.*

Determining whether the symbolic dynamics is complete (as is the case for sufficiently separated disks), pruned (for example, for touching or overlapping disks), or only a first coarse graining of the topology (as, for example, for smooth potentials with islands of stability) requires case-by-case investigation, a discussion we postpone to sect. 10.3 and chapter 11. For the time being we assume that the disks are sufficiently separated that there is no additional pruning beyond the prohibition of self-bounces.

If there are no restrictions on symbols, the symbolic dynamics is complete, and *all* binary sequences are admissible itineraries. As this type of symbolic dynamics pops up frequently, we list the shortest binary prime cycles in table **??**.

[exercise 10.2]

Inspecting the figure 10.3 we see that the relative ordering of regions with differing finite itineraries is a qualitative, topological property of the flow, so it makes sense to define a simple "canonical" representative partition which in a simple manner exhibits spatial ordering common to an entire class of topologically similar nonlinear flows.

in depth:
chapter 19, p. 320

**Table 10.1:** Prime cycles for the binary symbolic dynamics up to length 9.

| $n_p$ | $p$ | $n_p$ | $p$ | $n_p$ | $p$ | $n_p$ | $p$ | $n_p$ | $p$ |
|---|---|---|---|---|---|---|---|---|---|
| 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 | | |

## 10.2 Stretch and fold

Symbolic dynamics for $N$-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.

Suppose concentrations of certain chemical reactants worry you, or the variations in the Chicago temperature, humidity, pressure and winds affect your mood. All such properties 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 for keeping going is high - for example, we try to stir up some tar, and observe it come to a dead stop the moment we cease our labors - the dynamics tends to settle into a simple limiting state. However, as the resistance to change decreases - the tar is heated up and we are more vigorous in our stirring - the dynamics becomes unstable.

If a flow is locally unstable but globally bounded, any open ball of initial points will be stretched out and then folded back.

At this juncture we show how this works on the simplest example: unimodal mappings of the interval. The erudite reader should skim through this chapter and then take a more demanding path, via the Smale horseshoes of chapter 11. Uni-

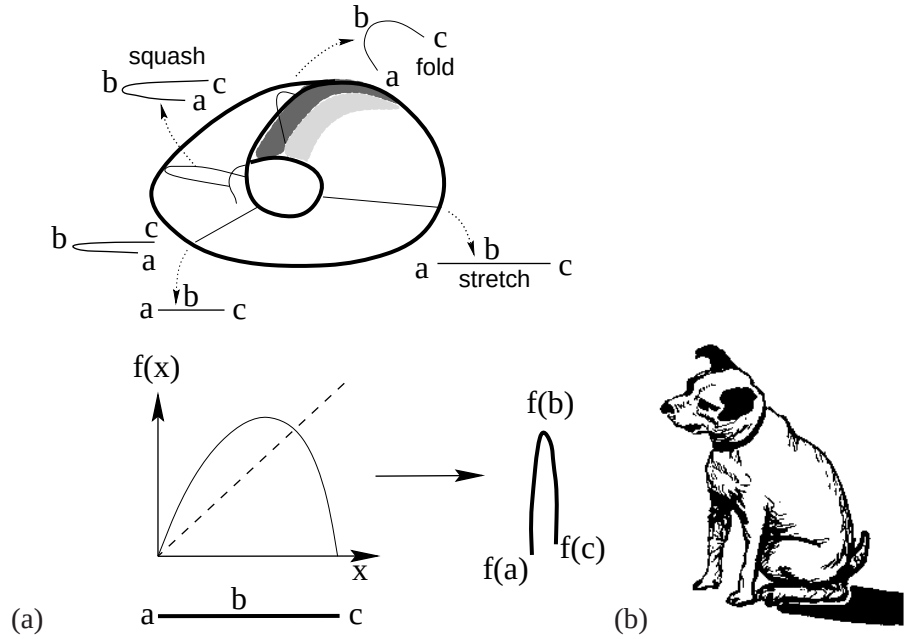



**Figure 10.5:** (a) A recurrent flow that stretches and folds. (b) The "stretch & fold" return map on the Poincaré section.

modal maps are easier, but physically less motivated. The Smale horseshoes are the high road, more complicated, but the right tool to generalize what we learned from 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.

### 10.2.1 Temporal ordering: itineraries

In this section we learn how to *name* (and, in chapter 13, how to *count*) periodic orbits for the simplest, and nevertheless very instructive case, for 1-dimensional maps of an interval.

Suppose that the compression of the folded interval in figure 10.5 is so fierce that we can neglect the thickness of the attractor. For example, the Rössler flow (2.17) is volume contracting, and an interval transverse to the attractor is stretched, folded and pressed back into a nearly 1-dimensional interval, typically compressed transversally by a factor of $\approx 10^{13}$ in one Poincaré section return. In such cases it makes sense to approximate the return map of a "stretch & fold" flow by a 1-dimensional map.

The simplest mapping of this type is *unimodal;* interval is stretched and folded only once, with at most two points mapping into a point in the refolded interval. A unimodal map $f(x)$ is a 1-dimensional function $\mathbb{R} \to \mathbb{R}$ defined on an interval $\mathcal{M} \in \mathbb{R}$ with a monotonically increasing (or decreasing) branch, a *critical point* (or interval) $x_c$ for which $f(x_c)$ attains the maximum (minimum) value, followed by a monotonically decreasing (increasing) branch. *Uni*-modal means that the map is a 1-humped map with one critical point within interval $\mathcal{M}$. A *multi*-modal map has several critical points within interval $\mathcal{M}$.

**Example 10.4** ***Complete tent map, quadratic map:*** *The simplest examples of*

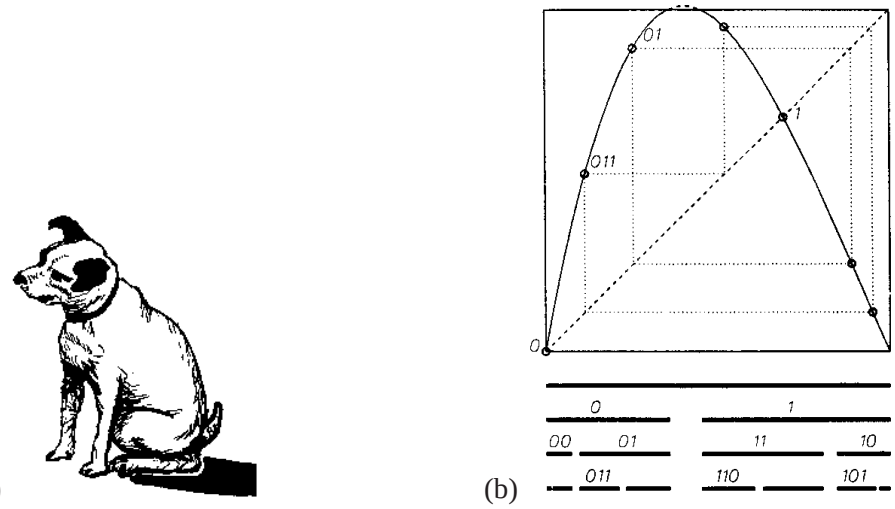



**Figure 10.6:** (a) The complete tent map together with intervals that follow the indicated itinerary for $n$ steps. (b) A unimodal repeller with the remaining intervals after 1, 2 and 3 iterations. Intervals marked $s_1 s_2 \cdots s_n$ are unions of all points that do not escape in $n$ iterations, and follow the itinerary $S^+ = s_1 s_2 \cdots s_n$. Note that the spatial ordering does not respect the binary ordering; for example $x_{00} < x_{01} < x_{11} < x_{10}$. Also indicated: the fixed points $x_0$, $x_1$, the 2-cycle $\overline{01}$, and the 3-cycle $\overline{011}$.

(a) (b)

*unimodal maps are the* complete tent map, *figure 10.6 (a),*

$$f(\gamma) = 1 - 2|\gamma - 1/2|\,, \tag{10.6}$$

*and the quadratic map (sometimes also called the* logistic map*)*

$$x_{t+1} = 1 - a x_t^2\,, \tag{10.7}$$

*with the one critical point at $x_c = 0$. Further examples are the repelling unimodal map of figure 10.6 (b) and the piecewise linear tent map (10.6).*

*Such dynamical systems are irreversible (the inverse of $f$ is double-valued), but, as we shall show in sect. 11.3, they may nevertheless serve as effective descriptions of invertible 2-dimensional hyperbolic flows.*

*For the unimodal maps of figure 10.6 a Markov partition of the unit interval $\mathcal{M}$ is given by the two intervals $\{\mathcal{M}_0, \mathcal{M}_1\}$. We refer to (10.6) as the "complete" tent map because its symbolic dynamics is complete binary: as both $f(\mathcal{M}_0)$ and $f(\mathcal{M}_1)$ fully cover $\mathcal{M}_0$ and $\mathcal{M}_1$, the corresponding transition matrix is a [2×2] matrix with all entries equal to 1, as in (10.3). As binary symbolic dynamics pops up frequently in applications, we list the shortest binary prime cycles in table* **??**.

**Example 10.5** ***Lorenz flow: a 1-d return map*** *We now deploy the symmetry of Lorenz flow to streamline and complete analysis of the Lorenz strange attractor commenced in example 9.2.*

*The dihedral $D_1 = \{e, R\}$ symmetry identifies the two equilibria $EQ_1$ and $EQ_2$, and the traditional "two-eared" Lorenz flow figure 2.4 is replaced by the "single-eared" flow of figure 9.2 (a). Furthermore, symmetry identifies two sides 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 9.2 (a).*

*Example 4.7 explained the genesis of the $x_{EQ1}$ equilibrium unstable manifold, its orientation and thickness, its collision with the $z$-axis, and its heteroclinic connection to the $x_{EQ0} = (0, 0, 0)$ equilibrium. All that remains is to describe how the $EQ_0$ neighborhood connects back to the $EQ_1$ unstable manifold. Figure 9.2 now shows clearly how the Lorenz dynamics is pieced together from the 2 equilibria and their unstable manifolds:*

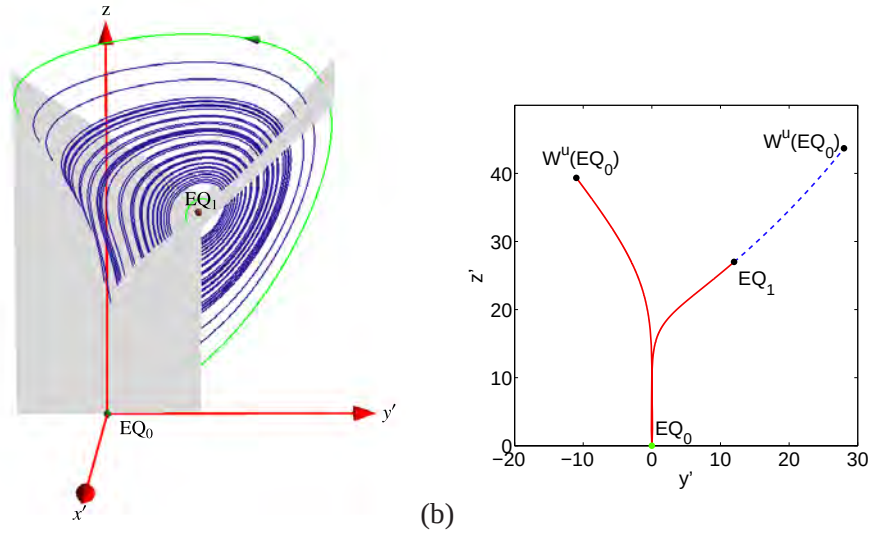



**Figure 10.7:** (a) A Poincaré section of the Lorer flow in the doubled-polar angle representation, fig ure 10.7, given by the $[y', z]$ plane that contains th $z$-axis and the equilibrium $EQ_1$. $x'$ axis points to ward the viewer. (b) The Poincaré section of th Lorenz flow by the section plane (a); compare wit figure 3.7. Crossings *into* the section are marke red (solid) and crossings *out of* the section a marked blue (dotted). Outermost points of bot in- and out-sections are given by the $EQ_0$ unstabl manifold $W^u(EQ_0)$ intersections. (E. Siminos)

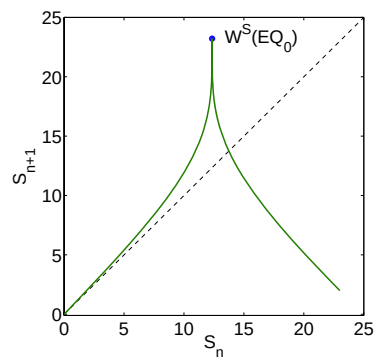



**Figure 10.8:** The Poincaré return map $s_{n+1} = P(s_n)$ parameterized by Euclidean arclength $s$ measured along the $EQ_1$ unstable manifold, from $x_{EQ1}$ to $W^u(EQ_0)$ section point, uppermost right point of the blue segment in figure 10.7 (b). The critical point (the "crease") of the map is 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)

*Having completed the descent to $EQ_0$, the infinitesimal neighborhood of the heteroclinic $EQ_1 \to EQ_0$ 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 at 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 the figure 10.7 (b), with the outermost edge following the unstable manifold of $EQ_0$.*

*Now the stage is set for construction of Poincaré sections and associated Poincaré return maps. There are two natural choices; the section at $EQ_0$, lower part of figure 10.7 (b), and the section (blue) above $EQ_1$. The first section, together with the blowup of the $EQ_0$ neighborhood, figure 4.7 (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 is given by figure 10.8.*

*The rest is straight sailing: to accuracy $10^{-4}$ the return map is unimodal, its "critical" point's forward trajectory yields the kneading sequence, and the admissible binary sequences, so any number of cycle points can be accurately determined from this 1-dimensional return map, and the 3-d cycles then verified by integrating the Lorenz differential equations (2.12). The map is everywhere expanding on the strange attractor, so it is no wonder mathematicians can here make the ergodicity rigorous.*

*Finally, the relation between the full state space periodic orbits, and the fundamental domain (9.16) reduced orbits: Full state space cycle pairs $p$, $Rp$ map into a single cycles $\tilde{p}$ in the fundamental domain, and any self-dual cycle $p = Rp = \tilde{p}R\tilde{p}$ is a*

*repeat of a relative periodic orbit $\tilde{p}$.*

*But there is trouble in paradise. By a fluke, the Lorenz attractor, the first flow to popularize strange attractors, turns to be topologically one of the simplest strange attractors. But it is not "uniformly hyperbolic." The flow near $EQ_1$ is barely unstable, while the flow near $EQ_0$ is arbitrarily unstable. So binary enumeration of cycles mixes cycles of vastly different stabilities, and is not very useful - presumably the practical way to compute averages is by stability ordering.*

*(E. Siminos and J. Halcrow)*

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 value $f(x_c)$ belongs neither to the left nor to the right partition $\mathcal{M}_i$, and is denoted by its own symbol $s = C$. As we shall see, its preimages serve as partition boundary points.

The trajectory $x_1, x_2, x_3, \ldots$ 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 (10.15) for a given trajectory,

$$s_n = \begin{cases} 1 & \text{if } x_n > x_c \\ 0 & \text{if } x_n < x_c \end{cases} . \tag{10.8}$$

We shall 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 corresponding *spatial* ordering of points that belong to a given trajectory?

### 10.2.2 Spatial ordering, 1-*d* maps

> Tired of being harassed by your professors? Finish, get a job, do combinatorics your own way, while you still know everything.
>
> —Professor Gatto Nero

Suppose you have succeeded in constructing a covering symbolic dynamics, such as for 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 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 11.8]

The easiest point of departure is to start out 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:

[section 10.3]

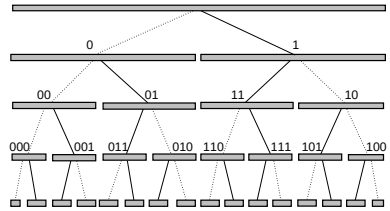



**Figure 10.9:** Alternating binary tree relates the itinerary labeling of the unimodal map figure 10.6 intervals to their spatial ordering. Dotted line stands for 0, full line for 1; the binary sub-tree whose root is a full line (symbol 1) reverses the orientation, due to the orientation reversing fold in figures 10.5 and 10.6.

The admissibility criterion eliminates *all* itineraries that cannot occur for a given unimodal map.

The tent map (10.6) consists of two straight segments joined at $x = 1/2$. The symbol $s_n$ defined in (10.8) equals 0 if the function increases, and 1 if the function decreases. The piecewise linearity of the map makes it possible to 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 to state the answer, and then hope that you will figure it out by yourself. Once you figure it out, feel free to complain that the way the rule is stated here is incomprehensible, and shows us how you did it better.

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 following algorithm:

$$
\begin{aligned}
w_{n+1} &= \begin{cases} w_n & \text{if } s_{n+1} = 0 \\ 1 - w_n & \text{if } s_{n+1} = 1 \end{cases}, \qquad w_1 = s_1 \\
\gamma(S^+) &= 0.w_1 w_2 w_3 \ldots = \sum_{n=1}^{\infty} w_n/2^n. \qquad (10.9)
\end{aligned}
$$

This follows by inspection from the binary tree of figure 10.9.

[exercise 10.4]

**Example 10.6** ***Converting*** $\gamma$ ***to*** $S^+$***:*** *$\gamma$ whose itinerary is $S^+ = 0110000\cdots$ is given by the binary number $\gamma = .010000\cdots$. Conversely, the itinerary of $\gamma = .01$ is $s_1 = 0$, $f(\gamma) = .1 \to s_2 = 1$, $f^2(\gamma) = f(.1) = 1 \to s_3 = 1$, etc..*

We shall refer to $\gamma(S^+)$ as the *(future) topological coordinate*. $w_t$'s are the digits in the binary expansion of the starting point $\gamma$ for the complete tent map (10.6). In the left half-interval the map $f(x)$ acts by multiplication by 2, while 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 $x_0 \to S^+(x_0) \to \gamma_0 = \gamma(S^+)$ is a *topological conjugacy* which maps the trajectory of an initial point $x_0$ under iteration of a given unimodal map to that initial point $\gamma$ for which the trajectory of the "canonical" unimodal map (10.6) has the same itinerary. The virtue of this conjugacy is that it *preserves the ordering* for any unimodal map in the sense that if $\overline{x} > x$, then $\overline{\gamma} > \gamma$.

**Figure 10.10:** The "dike" map obtained by slicing of a top portion of the tent map figure 10.6 (a). Any 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 tent map figure 10.6 (a) orbit.

## 10.3 Kneading theory

(K.T. Hansen and P. Cvitanović)

The main motivation for being mindful of spatial ordering of temporal itineraries is that this spatial ordering provides us with criteria that separate inadmissible orbits from those realizable by the dynamics. For 1-dimensional mappings the *kneading theory* provides such criterion of admissibility.

If the parameter in the quadratic map (10.7) is $a > 2$, then the iterates of the critical point $x_c$ diverge for $n \to \infty$. As long as $a \geq 2$, 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 $n \to \infty$ limit of the $n$th level covering intervals sketched in figure 10.6.

For $a < 2$ 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 $x_1 \to x_2 \to x_3 \to \ldots$ 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 the *kneading sequence* of the map. The corresponding topological coordinate is called the *kneading value*

$$\kappa = \gamma(K) = \gamma(S^+(x_c)). \tag{10.10}$$

A map with the same kneading sequence $K$ as $f(x)$, such as the dike map figure 10.10, is obtained by slicing off all $\gamma(S^+(x_0)) > \kappa$,

$$f(\gamma) = \begin{cases} f_0(\gamma) = 2\gamma & \gamma \in I_0 = [0, \kappa/2) \\ f_c(\gamma) = \kappa & \gamma \in I_c = [\kappa/2, 1 - \kappa/2] \\ f_1(\gamma) = 2(1 - \gamma) & \gamma \in I_1 = [1 - \kappa/2, 1] \end{cases} . \tag{10.11}$$

The dike map is the complete tent map figure 10.6 (a) with the top sliced off. It is convenient for coding the symbolic dynamics, as those $\gamma$ values that survive the

pruning are the same as for the complete tent map figure 10.6 (a), and are easily converted into admissible itineraries by (10.9).

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

$$\hat{\gamma}(S^+) = \sup_m \gamma(\sigma^m(S^+)) \tag{10.12}$$

be the *maximal value*, the highest topological coordinate reached by the orbit $x_1 \to x_2 \to x_3 \to \ldots$. We shall call the interval $(\kappa, 1]$ the *primary pruned interval*. The orbit $S^+$ is inadmissible if $\gamma$ of any shifted sequence of $S^+$ falls into this interval.

**Criterion of admissibility:** *Let $\kappa$ 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 $\hat{\gamma}(S^+) \leq \kappa$.*

While a unimodal map may depend on many arbitrarily chosen parameters, its dynamics determines the unique kneading value $\kappa$. We shall call $\kappa$ 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 $\kappa$ as a function of the map parameter such as $a$ in (10.7) correspond to inadmissible values of the topological parameter. Each jump in $\kappa$ corresponds to a stability window associated with a stable cycle of a smooth unimodal map. For the quadratic map (10.7) $\kappa$ increases monotonically with the parameter $a$, but for a general unimodal map such monotonicity need not hold.

For further details of unimodal dynamics, the reader is referred to appendixD.1. As we shall see in sect. 11.5, 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.

## 10.4 Markov graphs

### 10.4.1 Finite memory

In the complete $N$-ary symbolic dynamics case (see example (10.3)) the choice of the next symbol requires no memory of the previous ones. However, any further refinement of the partition requires finite memory.

For example, 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 10.4, as well as figure 1.9. Topologically $f$ acts as a left shift (11.10), and its action on the rectangle [.01] is to move the decimal point to the right, to

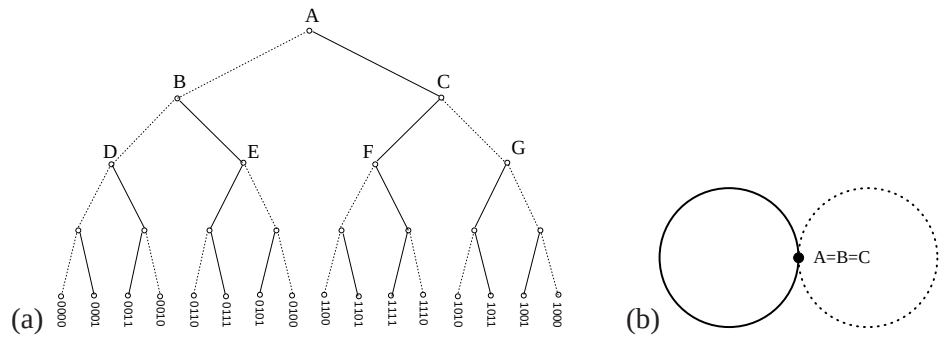



**Figure 10.11:** (a) The self-similarity of the complete binary symbolic dynamics represented by a binary tree (b) identification of nodes $B = A$, $C = A$ leads to the finite 1-node, 2-links Markov graph. All admissible itineraries are generated as walks on this finite Markov graph.

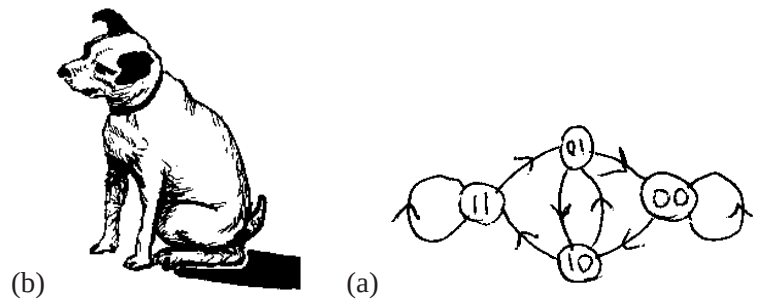



**Figure 10.12:** (a) The 2-step memory Markov graph, links version obtained by identifying nodes $A = D = E = F = G$ in figure 10.11 (a). Links of this graph correspond to the matrix entries in the transition matrix (10.13). (b) the 2-step memory Markov graph, node version.

[0.1], forget the past, [.1], and land in either of the two rectangles $\{[.10], [.11]\}$. Filling in the matrix elements for the other three initial states we obtain the 1-step memory transition matrix acting on the 4-state vector

[exercise 10.7]

$$\phi' = T\phi \;=\; \begin{pmatrix} T_{00,00} & 0 & T_{00,10} & 0 \\ T_{01,00} & 0 & T_{01,10} & 0 \\ 0 & T_{10,01} & 0 & T_{10,11} \\ 0 & T_{11,01} & 0 & T_{11,11} \end{pmatrix} \begin{pmatrix} \phi_{00} \\ \phi_{01} \\ \phi_{10} \\ \phi_{11} \end{pmatrix}. \tag{10.13}$$

By the same token, for $M$-step memory the only nonvanishing matrix elements are of the form $T_{s_1 s_2 \ldots s_{M+1}, s_0 s_1 \ldots s_M}$, $s_{M+1} \in \{0, 1\}$. This is a sparse matrix, as the only non vanishing entries in the $m = s_0 s_1 \ldots s_M$ column of $T_{dm}$ are in the rows $d = s_1 \ldots s_M 0$ and $d = s_1 \ldots s_M 1$. If we increase the number of steps remembered, the transition matrix grows big 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 representation is afforded by Markov graphs, which are not only compact, but also give us an intuitive picture of the topological dynamics.

[exercise 13.1]

Construction of a good Markov graph is, like combinatorics, unexplainable. The only way to learn is by some diagrammatic gymnastics, so we work our way through a sequence of exercises in lieu of plethora of baffling definitions.

[exercise 13.4]

[exercise 13.1]

To start with, what do finite graphs have to do with infinitely long trajectories? To understand the main idea, let us construct a graph that enumerates all possible itineraries for the case of complete binary symbolic dynamics.

Mark a dot "·" on a piece of paper. Draw two short 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 signifying "0". Repeat the procedure for each of the

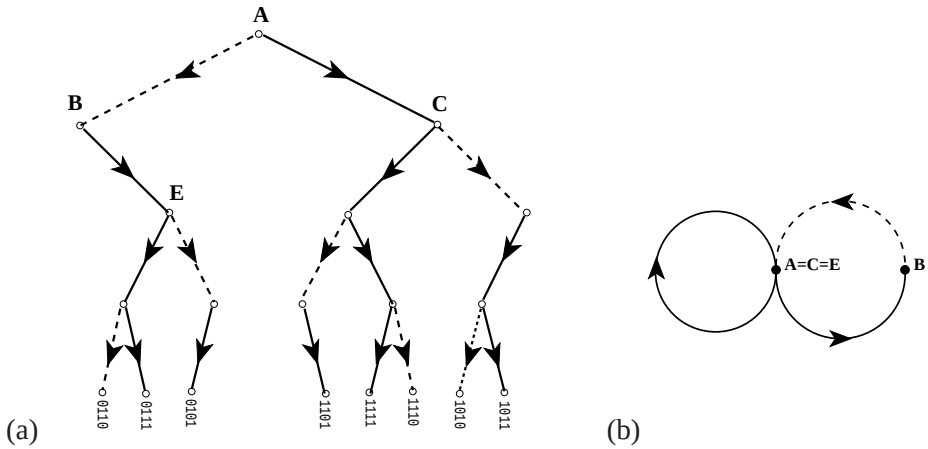



**Figure 10.13:** (a) The self-similarity of the _00_ pruned binary tree: trees originating from nodes $C$ and $E$ are the same as the entire tree. (b) Identification of nodes $A = C = E$ leads to the finite 2-node, 3-links Markov graph; as 0 is always followed by 1, the walks on this graph generate only the admissible itineraries.

two new dots, and then for the four dots, and so on. The result is the binary tree of figure 10.11 (a). Starting at the top node, the tree enumerates exhaustively all distinct finite itineraries

$$\{0, 1\},$$
$$\{00, 01, 10, 11\},$$
$$\{000, 001, 010, \cdots\}, \cdots .$$

The $M = 4$ nodes in figure 10.11 (a) correspond to the 16 distinct binary strings of length 4, and so on. By habit we have drawn the tree as the alternating binary tree of figure 10.9, but that has no significance as far as enumeration of itineraries is concerned - an ordinary binary tree 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, and can be identified. To say it 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 Markov graph of figure 10.11 (b): any itinerary generated by the binary tree figure 10.11 (a), 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 Markov graphs can do the job as well, and might be sometimes preferable. For example, identifying the trees originating in $D$, $E$, $F$ and $G$ with the entire tree leads to the 2-step memory Markov graph of figure 10.12a. The corresponding transition matrix is given by (10.13).

in depth:
chapter 11, p. 174

fast track:
chapter 13, p. 212

## 10.5 Symbolic dynamics, basic notions

In this section we collect the basic notions and definitions of symbolic dynamics. The reader might prefer to skim through this material on first reading, 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 $x_1, x_2, x_3, \ldots$ of the initial point $x_0$ is generated by

$$x_{n+1} = f(x_n)\,, \tag{10.14}$$

then the itinerary is given by the symbol sequence

$$s_n = s \qquad \text{if} \qquad x_n \in \mathcal{M}_s F\,. \tag{10.15}$$

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 to the point $x_0$. To each point $x_0$ in the dynamical space we thus associate a bi-infinite itinerary

$$S(x_0) = (s_k)_{k\in\mathbb{Z}} = S^-.S^+ = \cdots s_{-2} s_{-1} s_0 . s_1 s_2 s_3 \cdots\,. \tag{10.16}$$

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*

$$\mathcal{A}^{\mathbb{Z}} = \{(s_k)_{k\in\mathbb{Z}} : s_k \in \mathcal{A} \text{ for all k } \in \mathbb{Z}\}\,. \tag{10.17}$$

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.

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 (10.15), a forward iteration $x \to x' = f(x)$ shifts the entire itinerary to the left through the "decimal point." This operation, denoted by the shift operator $\sigma$,

$$\sigma(\cdots s_{-2} s_{-1} s_0 . s_1 s_2 s_3 \cdots) = \cdots s_{-2} s_{-1} s_0 s_1 . s_2 s_3 \cdots\,, \tag{10.18}$$

demoting the current partition label $s_1$ from the future $S^+$ to the "has been" itinerary $S^-$. The inverse shift $\sigma^{-1}$ 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 $n_b$. A state space trajectory is *periodic* if it returns to its initial point after a finite time; in the shift space the trajectory is periodic if its itinerary is an infinitely repeating block $p^\infty$. We shall refer to the set of periodic points that belong to a given periodic orbit as a *cycle*

$$p = \overline{s_1 s_2 \cdots s_{n_p}} = \{x_{s_1 s_2 \cdots s_{n_p}}, x_{s_2 \cdots s_{n_p} s_1}, \cdots, x_{s_{n_p} s_1 \cdots s_{n_p-1}}\}\,. \tag{10.19}$$

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 infinitely repeating basic block; we shall omit the bar whenever it is clear from the context that the trajectory is periodic. Each *cycle point* is labeled by the first $n_p$ steps of its future itinerary. For example, the 2nd cycle point is labeled by

$$x_{s_2 \cdots s_{n_p} s_1} = x_{\overline{s_2 \cdots s_{n_p} s_1}.\overline{s_2 \cdots s_{n_p} s_1}}\,.$$

A *prime* cycle $p$ of length $n_p$ 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 literature such cycle is sometimes called *primitive*; we shall refer to it as "prime" throughout this text).

**Partitions.** A partition is called *generating* if every infinite symbol sequence corresponds to a distinct point in the state space. Finite Markov partition (10.4) is an example. Constructing a generating partition for a given system is a difficult problem. In examples to follow we shall concentrate on cases which allow finite partitions, but in practice almost any generating partition of interest is infinite.

A mapping $f : \mathcal{M} \to \mathcal{M}$ together with a partition $\mathcal{A}$ induces *topological dynamics* $(\Sigma, \sigma)$, where the *subshift*

$$\Sigma = \{(s_k)_{k\in\mathbb{Z}}\}\,, \tag{10.20}$$

is the set of all *admissible* infinite itineraries, and $\sigma : \Sigma \to \Sigma$ is the shift operator (10.18). The designation "subshift" comes form the fact that $\Sigma \subset \mathcal{A}^{\mathbb{Z}}$ is the subset of the full shift (10.17). One of our principal tasks in developing symbolic dynamics of dynamical systems that occur in nature will be to determine $\Sigma$, the set of all bi-infinite itineraries $S$ that are actually realized by the given dynamical system.

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 $\cdots s_{-2} s_{-1} s_0 . s_1 s_2 s_3 \cdots$ to each distinct trajectory, but there might exist full shift symbol sequences (10.17) which are not realized as trajectories; such sequences are called *inadmissible*, and we say that the symbolic dynamics is *pruned*. The

word is suggested by "pruning" of branches corresponding to forbidden sequences for symbolic dynamics organized hierarchically into a tree structure, as explained in sect. 10.4.

**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,

$$\mathcal{G} = \{b_1, b_2, \cdots b_k\} \ , \tag{10.21}$$

where a *pruning block* $b$ is a sequence of symbols $b = s_1 s_2 \cdots s_{n_b}$, $s \in \mathcal{A}$, of finite length $n_b$. In this case we can always construct a finite Markov partition (10.4) 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 $M$. In the new alphabet the grammar rules are implemented by setting $T_{ij} = 0$ in (10.3) for forbidden transitions.

A topological dynamical system $(\Sigma, \sigma)$ for which all admissible itineraries are generated by a finite transition matrix

$$\Sigma = \{(s_k)_{k\in\mathbb{Z}} \ : \ T_{s_k s_{k+1}} = 1 \quad \text{for all } k\} \tag{10.22}$$

is called a subshift of *finite type*. Such systems are particularly easy to handle; the topology can be converted into symbolic dynamics by representing the transition matrix by a finite directed *Markov graph*, a convenient visualization of topological dynamics.

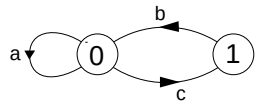
**Markov graphs.** A Markov graph describes compactly the ways in which the state space regions map into each other, accounts for finite memory effects in dynamics, and generates the totality of admissible trajectories as the set of all possible walks along its links.

A Markov graph consists of a set of *nodes* (or *vertices*, or *states*), one for each state in the alphabet $\mathcal{A} = \{A, B, C, \cdots, Z\}$, connected by a set of directed *links* (*edges*, *arcs*). Node $i$ is connected by a directed link to node $j$ whenever the transition matrix element (10.2) takes value $T_{ij} = 1$. There might be a set of links connecting two nodes, or links that originate and terminate on the same node. Two graphs are isomorphic if one can be obtained from the other by relabeling links and nodes; for us they are one and the same graph. As we are interested in recurrent 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.

**Example 10.7** ***"Golden mean" pruning*** *Consider a simple subshift on two-state partition $\mathcal{A} = \{0, 1\}$, with the simplest grammar $\mathcal{G}$ possible: a single pruning block $b =$*

**Figure 10.14:** (a) The transition matrix for binary alphabet $\mathcal{A} = \{0, 1\}$, $b$ = _11_ pruned. (b) The corresponding Markov graph.

$$T = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \text{(b)}$$




*_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 for this grammar is given in figure 10.14 (a). The corresponding finite 2-node, 3-links Markov graph, with nodes coding the symbols, is given in figure 10.14 (b). All admissible itineraries are generated as walks on this finite Markov graph.*

in depth:
chapter 11, p. 174

## Résumé

In chapters 16 and 17 we will establish that spectra of evolution operators can be extracted from periodic orbit sums:

$$\sum \text{(spectral eigenvalues)} = \sum \text{(periodic orbits)} \ .$$

In order to implement this theory we need to know what periodic orbits can exist, and the symbolic dynamics developed above and in chapter 11 is an invaluable tool toward this end.

## Commentary

**Remark 10.1** Symbolic dynamics, history and good taste. For a brief history of symbolic dynamics, from J. Hadamard in 1898 onward, see Notes to chapter 1 of Kitchens monograph [1], a very clear and enjoyable mathematical introduction to topics discussed here. Diacu and Holmes [2] provide an excellent survey of symbolic dynamics applied to of celestial mechanics. Finite Markov graphs or finite automata are discussed in refs. [3, 4, 5, 6]. They belong to the category of regular languages. A good hands-on introduction to symbolic dynamics is given in ref. [12].

The binary labeling of the once-folding map periodic points was introduced by Myrberg [13] for 1-dimensional maps, and its utility to 2-dimensional maps has been emphasized in refs. [8, 12]. For 1-dimensional maps it is now customary to use the *R*-*L* notation of Metropolis, Stein and Stein [14, 15], indicating that the point $x_n$ lies either to the left or to the right of the critical point in figure 10.6. The symbolic dynamics of such mappings has been extensively studied by means of the Smale horseshoes, see for example ref. [16]. Using letters rather than numerals in symbol dynamics alphabets probably reflects good taste. We prefer numerals for their computational convenience, as they speed up the implementation of conversions into the topological coordinates $(\delta, \gamma)$ introduced in sect. 11.4.1. The alternating binary ordering of figure 10.9 is related to the Gray codes of computer science [12].

**Remark 10.2** Counting prime cycles. Duval has an efficient algorithm for generating Lyndon words (non-periodic necklaces, i.e., prime cycle itineraries).

**Remark 10.3** Inflating Markov graphs. In the above examples the symbolic dynamics has been encoded by labeling links in the Markov graph. Alternatively one can encode the dynamics by labeling the nodes, as in figure 10.12, where the 4 nodes refer to 4 Markov partition regions $\{\mathcal{M}_{00}, \mathcal{M}_{01}, \mathcal{M}_{10}, \mathcal{M}_{11}\}$, and the 8 links to the 8 non-zero entries in the 2-step memory transition matrix (10.13).

fast track:
chapter 13, p. 212

## Exercises

10.1. **Binary symbolic dynamics.** Verify that the shortest prime binary cycles of the unimodal repeller of figure 10.6 are $\overline{0}$, $\overline{1}$, $\overline{01}$, $\overline{001}$, $\overline{011}$, $\cdots$. Compare with table **??**. Try to sketch them in the graph of the unimodal function $f(x)$; compare ordering of the periodic points with figure 10.9. The point is that while overlayed on each other the longer cycles look like a hopeless jumble, the cycle points are clearly and logically ordered by the alternating binary tree.

10.2. **Generating prime cycles.** Write a program that generates all binary prime cycles up to given finite length.

10.3. **A contracting baker's map.** Consider a contracting (or "dissipative") baker's defined in exercise 4.6.

The symbolic dynamics encoding of trajectories is realized via symbols 0 ($y \le 1/2$) and 1 ($y > 1/2$). Consider the observable $a(x, y) = x$. Verify that for any periodic orbit $p$ $(\epsilon_1 \ldots \epsilon_{n_p})$, $\epsilon_i \in \{0, 1\}$

$$A_p = \frac{3}{4} \sum_{j=1}^{n_p} \delta_{j,1} \,.$$

10.4. **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 (10.9). This follows by inspection from the binary tree of figure 10.9.

10.5. **Unimodal map kneading value.** Consider the 1-$d$ quadratic map

$$f(x) = Ax(1-x)\,, \qquad A = 3.8\,. \tag{10.23}$$

(a) (easy) Plot (10.23), 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 10.6 (b). Note, however, that some of the intervals of figure 10.6 (b) do not appear in this case - they are *pruned*.

(c) (medium) Produce ChaosBook.org quality figure 10.6 (a).

(d) (easy) Check numerically that $K = S^+(x_c)$, the itinerary or the "kneading sequence" of the critical point is

$$K = 1011011110110111101011110111110\ldots$$

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 (10.9),

$$w_{n+1} = \begin{cases} w_n & \text{if } s_{n+1} = 0 \\ 1 - w_n & \text{if } s_{n+1} = 1 \end{cases}, \qquad w_1 =$$

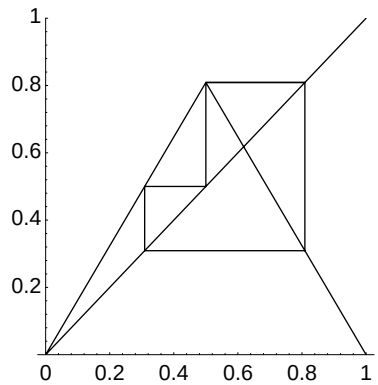
$$\gamma(S^+) = 0.w_1 w_2 w_3 \ldots = \sum_{n=1}^{\infty} w_n / 2^n \,.$$

(e) (medium) List the the corresponding kneading value (10.10) sequence $\kappa = \gamma(K)$ to the same number of digits as $K$.

(f) (hard) Plot the missing dike map, figure 10.10, in ChaosBook.org quality, 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 complete tent map figure 10.6 (a), see (10.11).

How this kneading sequence is converted into a series of pruning rules is a dark art, relegated to sect. 13.6.

10.6. **"Golden mean" pruned map.** Consider a symmetrical tent map on the unit interval such that its highest point belongs to a 3-cycle:




(a) Find the absolute value $\Lambda$ for the slope (the two different slopes $\pm\Lambda$ just differ by a sign) where the maximum at $1/2$ is part of a period three orbit, as in the figure.

(b) Show that no orbit of this map can visit the region $x > (1+\sqrt{5})/4$ more than once. Verify that once an orbit exceeds $x > (\sqrt{5}-1)/4$, it does not reenter the region $x < (\sqrt{5}-1)/4$.

(c) If an orbit is in the interval $(\sqrt{5}-1)/4 < x < 1/2$, where will it be on the next iteration?

(d) 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.

(e) On the second thought, is there a periodic orbit that violates the above _00_ pruning rule?

For continuation, see exercise 13.6 and exercise 17.2. See also exercise 13.7 and exercise 13.8.

10.7. **Binary 3-step transition matrix.** Construct [8×8] binary 3-step transition matrix analogous to the 2-step transition matrix (10.13). Convince yourself that the number of terms of contributing to $\mathrm{tr}\, T^n$ is independent of the memory length, and that this $[2^m \times 2^m]$ trace is well defined in the infinite memory limit $m \to \infty$.

## References

[10.1] B.P. Kitchens, Symbolic dynamics: one-sided, two-sided, and countable state Markov shifts (Springer, Berlin 1998).

[10.2] F. Diacu and P. Holmes, *Celestial Encounters, The Origins of Chaos and Stability* (Princeton Univ. Press, Princeton NJ 1996).

[10.3] A. Salomaa, *Formal Languages* (Academic Press, San Diego, 1973).

[10.4] J.E. Hopcroft and J.D. Ullman, *Introduction to Automata Theory, Languages, and Computation* (Addison-Wesley, Reading MA, 1979).

[10.5] D.M. Cvetković, M. Doob and H. Sachs, *Spectra of Graphs* (Academic Press, New York, 1980).

[10.6] P. Grassberger, "On the symbolic dynamics of the one-humped map of the interval" *Z. Naturforsch. A* **43**, 671 (1988).

[10.7] P. Grassberger, R. Badii and A. Politi, *Scaling laws for invariant measures on hyperbolic and nonhyperbolic attractors*, *J. Stat. Phys.* **51**, 135 (1988).

[10.8] S. Isola and A. Politi, "Universal encoding for unimodal maps," *J. Stat. Phys.* **61**, 259 (1990).

[10.9] Y. Wang and H. Xie, "Grammatical complexity of unimodal maps with eventually periodic kneading sequences," *Nonlinearity* **7**, 1419 (1994).

[10.10] A. Boyarski, M. Skarowsky, *Trans. Am. Math. Soc.* **225**, 243 (1979); A. Boyarski, *J.Stat. Phys.* **50**, 213 (1988).

[10.11] C.S. Hsu, M.C. Kim, *Phys. Rev. A* **31**, 3253 (1985); N. Balmforth, E.A. Spiegel, C. Tresser, *Phys. Rev. Lett.* **72**, 80 (1994).

[10.12] D.A. Lind and B. Marcus, *An introduction to symbolic dynamics and coding* (Cambridge Univ. Press, Cambridge 1995).

[10.13] P.J. Myrberg, *Ann. Acad. Sc. Fenn., Ser. A,* **256**, 1 (1958); **259**, 1 (1958).

[10.14] N. Metropolis, M.L. Stein and P.R. Stein, "On Finite Limit Sets for Transformations on the Unit Interval," *J. Comb. Theo.* **15**, 25 (1973).

[10.15] P. Collet and J.P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems* (Birkhauser, Boston, 1980).

[10.16] J. Guckenheimer and P. Holmes, *Non-linear Oscillations, Dynamical Systems and Bifurcations of Vector Fields* (Springer, New York, 1986).

[10.17] R.L. Devaney, *An Introduction to Chaotic Dynamical Systems* (Addison-Wesley, Reading MA, 1987).

[10.18] R.L. Devaney, *A First Course in Chaotic Dynamical Systems* (Addison-Wesley, Reading MA, 1992).

[10.19] Bai-Lin Hao, *Elementary symbolic dynamics and chaos in dissipative systems* (World Scientific, Singapore, 1989).

[10.20] E. Aurell, "Convergence of dynamical zeta functions," *J. Stat. Phys.* **58**, 967 (1990).

[10.21] M.J. Feigenbaum, *J. Stat. Phys.* **46**, 919 (1987); **46**, 925 (1987).

[10.22] P. Cvitanović, "Chaos for cyclists," in E. Moss, ed., *Noise and chaos in nonlinear dynamical systems* (Cambridge Univ. Press, Cambridge 1989).

[10.23] P. Cvitanović, "The power of chaos," in J.H. Kim and J. Stringer, eds., *Applied Chaos*, (John Wiley & Sons, New York 1992).

[10.24] P. Cvitanović, ed., *Periodic Orbit Theory - theme issue*, *CHAOS* **2**, 1-158 (1992).

[10.25] P. Cvitanović, "Dynamical averaging in terms of periodic orbits," *Physica* **D 83**, 109 (1995).

# Chapter 11

# Qualitative dynamics, for cyclists

> **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 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*

In sects. 14.1 and 10.1 we introduced the concept of partitioning the state space, in any way you please. In chapter 5 we established that stability eigenvalues of periodic orbits are invariants of a given flow. The invariance of stabilities of a periodic orbit is a local property of the flow.

For the Rössler flow of example 3.4, we have learned that the attractor is very thin, but otherwise the return maps that we found were disquieting – figure 3.6 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 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. 12.1), 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.

In this chapter we show that every equilibrium point and every periodic orbit carries with it stable and unstable manifolds which provide a topologically invariant *global* foliation of the state space. This qualitative dynamics of stretching

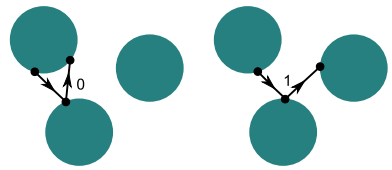



**Figure 11.1:** Binary labeling of trajectories of the symmetric 3-disk pinball; a bounce in which the trajectory returns to the preceding disk is labeled 0, and a bounce which results in continuation to the third disk is labeled 1.

and mixing enables us to partition the state space and assign symbolic dynamics itineraries to trajectories.

Given an itinerary, the topology of stretching and folding fixes the relative spatial ordering of trajectories, and separates the admissible and inadmissible itineraries. The level is distinctly cyclist, in distinction to the pedestrian tempo of the preceding chapter. Skip this chapter unless you really need to get into nitty-gritty details of symbolic dynamics.

fast track:
chapter 13, p. 212

## 11.1 Recoding, symmetries, tilings

In chapter 9 we made a claim that if there is a symmetry of dynamics, we must use it. So let's take the old pinball game and "quotient the state space by the symmetry or "desymmetrize."

Though a useful tool, Markov partitioning is not without drawbacks. One glaring shortcoming is that Markov partitions are not unique: any of many different partitions might do the job. The 3-disk system offers a simple illustration of different Markov partitioning strategies for the same dynamical system.

The $\mathcal{A} = \{1, 2, 3\}$ symbolic dynamics for 3-disk system is neither unique, nor necessarily the smartest one - before proceeding it pays to exploit the symmetries of the pinball in order to obtain a more efficient description. In chapter19 we shall be handsomely rewarded for our labors.

As the three disks are equidistantly spaced, our game of pinball has a sixfold symmetry. For instance, the cycles $\overline{12}$, $\overline{23}$, and $\overline{13}$ are related to each other by rotation by $\pm 2\pi/3$ or, equivalently, by a relabeling of the disks. 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 the collision. For the 3-disk game of pinball there are two topologically distinct kinds of collisions, figure 11.1:

[exercise 10.1]
[exercise 9.1]

$$s_i = \begin{cases} 0 & : \quad \text{pinball returns to the disk it came from} \\ 1 & : \quad \text{pinball continues to the third disk}\,. \end{cases} \tag{11.1}$$

This *binary* symbolic dynamics has two immediate advantages over the ternary one; the prohibition of self-bounces is automatic, and the coding utilizes the sym-

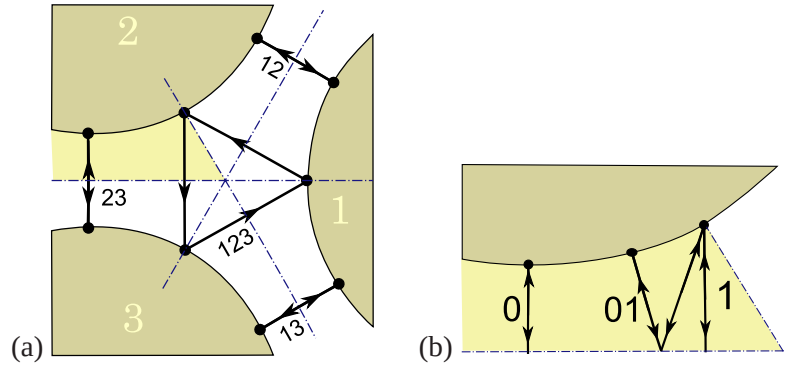



**Figure 11.2:** The 3-disk game of pinball with the disk radius : center separation ratio a:R = 1:2.5. (a) 2-cycles $\overline{12}$, $\overline{13}$, $\overline{23}$, and 3-cycles $\overline{123}$ and $\overline{132}$ (not drawn). (b) The fundamental domain, i.e., the small 1/6th wedge indicated in (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 9.4 for cycle $\overline{10}$ and further examples.

metry of the 3-disk pinball game in elegant manner. 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 **??**) are admissible itineraries. [exercise 10.2]

**Example 11.1** ***Recoding ternary symbolic dynamics in binary:*** *Given a ternary sequence and labels of 2 preceding disks, rule (11.1) fixes the subsequent binary symbols. Here we list an arbitrary ternary itinerary, and the corresponding binary sequence:*

$$\begin{array}{rcl} \textit{ternary} & : & 3\ 1\ 2\ 1\ 3\ 1\ 2\ 3\ 2\ 1\ 2\ 3\ 1\ 3\ 2\ 3 \\ \textit{binary} & : & \cdot\ 1\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0 \end{array} \tag{11.2}$$

*The first 2 disks initialize the trajectory and its direction;* $3 \mapsto 1 \mapsto 2 \mapsto \cdots$. *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, the same size partitions, and are coded by a single binary sequence. For periodic orbits, the equivalent ternary cycles reduce to binary cycles of 1/3, 1/2 or the same length. How this works is best understood by inspection of table* **??**, *figure 11.2 and figure 9.5.*

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 reflecting mirrors, see figure 11.2 (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 symbols block corresponds either to the full periodic orbit or to a relative periodic orbit (examples are shown in figure 11.2 and table **??**). An irreducible segment corresponds to a periodic orbit in the fundamental domain. Table **??** 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 19, 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 11.2** $C_2$ ***recoded:*** *As the simplest example of implementing the above scheme consider the* $C_2$ *symmetry of example 9.4. 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* $C_2$ *symmetric. The* $C_2$ *symmetry cycles separate into two classes, the self-dual configurations* $+-$, $++--$, $+++---$, $+--++-++-$, $\cdots$, *with multiplicity* $m_p = 1$, *and* [exercise 9.2]

**Table 11.1:** $C_{3v}$ correspondence between the binary labeled fundamental domain prime cycles $\tilde{p}$ and the full 3-disk ternary labeled cycles $p$, together with the $C_{3v}$ transformation that maps the end point of the $\tilde{p}$ cycle into the irreducible segment of the $p$ cycle, see sect. 9.3. Breaks in the above ternary sequences mark repeats of the irreducible segment. The multiplicity of $p$ cycle is $m_p = 6n_{\tilde{p}}/n_p$. The shortest pair of the fundamental domain cycles related by time reversal (but no spatial symmetry) are the 6-cycles $\overline{001011}$ and $\overline{001101}$.

| $\tilde{p}$ | $p$ | $\mathbf{g}_{\tilde{p}}$ | $\tilde{p}$ | $p$ | $\mathbf{g}_{\tilde{p}}$ |
|---|---|---|---|---|---|
| 0 | 1 2 | $\sigma_{12}$ | 000001 | 121212 131313 | $\sigma_{23}$ |
| 1 | 1 2 3 | $C$ | 000011 | 121212 313131 232323 | $C^2$ |
| 01 | 12 13 | $\sigma_{23}$ | 000101 | 121213 | $e$ |
| 001 | 121 232 313 | $C$ | 000111 | 121213 212123 | $\sigma_{12}$ |
| 011 | 121 323 | $\sigma_{13}$ | 001011 | 121232 131323 | $\sigma_{23}$ |
| 0001 | 1212 1313 | $\sigma_{23}$ | 001101 | 121231 323213 | $\sigma_{13}$ |
| 0011 | 1212 3131 2323 | $C^2$ | 001111 | 121231 232312 313123 | $C$ |
| 0111 | 1213 2123 | $\sigma_{12}$ | 010111 | 121312 313231 232123 | $C^2$ |
| 00001 | 12121 23232 31313 | $C$ | 011111 | 121321 323123 | $\sigma_{13}$ |
| 00011 | 12121 32323 | $\sigma_{13}$ | 0000001 | 1212121 2323232 3131313 | $C$ |
| 00101 | 12123 21213 | $\sigma_{12}$ | 0000011 | 1212121 3232323 | $\sigma_{13}$ |
| 00111 | 12123 | $e$ | 0000101 | 1212123 2121213 | $\sigma_{12}$ |
| 01011 | 12131 23212 31323 | $C$ | 0000111 | 1212123 | $e$ |
| 01111 | 12132 13123 | $\sigma_{23}$ | $\cdots$ | $\cdots$ | $\cdots$ |

*the asymmetric pairs $+, -, ++-, --+, \cdots$, with multiplicity $m_p = 2$. For example, as there is no absolute distinction between the "up" and the "down" spins, or the "left" or the "right" lobe, $\Lambda_+ = \Lambda_-$, $\Lambda_{++-} = \Lambda_{+--}$, and so on.*

[exercise 19.4]

*The symmetry reduced labeling $\rho_i \in \{0, 1\}$ is related to the standard $s_i \in \{+, -\}$ Ising spin labeling by*

$$\begin{aligned} &\text{If} \quad s_i = s_{i-1} \quad \text{then} \quad \rho_i = 1 \\ &\text{If} \quad s_i \neq s_{i-1} \quad \text{then} \quad \rho_i = 0 \end{aligned} \tag{11.3}$$

*For example, $\overline{+} = \cdots + + + + \cdots$ maps into $\cdots 111 \cdots = \overline{1}$ (and so does $\overline{-}$), $\overline{-+} = \cdots - + - + \cdots$ maps into $\cdots 000 \cdots = \overline{0}$, $\overline{-++-} = \cdots - - + + - - + + \cdots$ maps into $\cdots 0101 \cdots = \overline{01}$, and so forth. A list of such reductions is given in table* **??**.

**Example 11.3** $C_{3v}$ ***recoded - 3-disk game of pinball:***

*The $C_{3v}$ recoding can be worked out by a glance at figure 11.2 (a) (continuation of example 9.5). 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 11.2 (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, $\overline{23} = \cdots 232323 \cdots$ maps into $\cdots 000 \cdots = \overline{0}$ (and so do $\overline{12}$ and $\overline{13}$), $\overline{123} = \cdots 12312 \cdots$ maps into $\cdots 111 \cdots = \overline{1}$ (and so does $\overline{132}$), and so forth. A list of such reductions for short cycles is given in table* **??**, *figure 11.2 and figure 9.5.*

**Table 11.2:** Correspondence between the $C_2$ symmetry reduced cycles $\tilde{p}$ and the standard Ising model periodic configurations $p$, together with their multiplicities $m_p$. Also listed are the two shortest cycles (length 6) related by time reversal, but distinct under $C_2$.

| $\tilde{p}$ | $p$ | $m_p$ |
|---|---|---|
| 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 |

## 11.2 Going global: Stable/unstable manifolds

In the linear approximation, the fundamental matrix $M^t$ describes the shearing of an infinitesimal neighborhood in after a finite time $t$. Its eigenvalues and eigendirections describe deformation of an initial infinitesimal sphere of neighboring trajectories into an ellipsoid time $t$ later. Nearby trajectories separate exponentially along the unstable directions, approach each other along the stable directions, and maintain their distance along the marginal directions.

The fixed or periodic point $x^*$ fundamental matrix $M_p(x^*)$ eigenvectors (5.12) form a rectilinear coordinate frame in which the flow into, out of, or encircling the fixed point is linear in the sense of sect. 4.2. These eigendirections are numerically continued into global curvilinear invariant manifolds as follows.

The global continuations of the local stable, unstable eigendirections are called the *stable*, respectively *unstable manifolds*. They consist of all points which march into the fixed point forward, respectively backward in time

$$\begin{aligned} W^s &= \left\{x \in \mathcal{M} : f^t(x) - x^* \to 0 \text{ as } t \to \infty\right\} \\ W^u &= \left\{x \in \mathcal{M} : f^{-t}(x) - x^* \to 0 \text{ as } t \to \infty\right\} . \end{aligned} \tag{11.4}$$

The stable/unstable manifolds of a flow are rather hard to visualize, so as long as we are not worried about a global property such as the number of times they wind around a periodic trajectory before completing a par-course, we might just as well look at their Poincaré section return maps. Stable, unstable manifolds for maps are defined by

$$\begin{aligned} W^s &= \{x \in \mathcal{P} : f^n(x) - x^* \to 0 \text{ as } n \to \infty\} \\ W^u &= \{x \in \mathcal{P} : f^{-n}(x) - x^* \to 0 \text{ as } n \to \infty\} . \end{aligned} \tag{11.5}$$

Eigenvectors (real or complex pairs) of fundamental matrix $M_p(x^*)$ play a special role - on them the action of the dynamics is the linear multiplication by $\Lambda_i$ (for a real eigenvector) along 1-$d$ invariant curve $W^{u,s}_{(i)}$ or spiral in/out action in a 2-$D$ surface (for a complex pair). For $n \to \infty$ a finite segment on $W^s_{(e)}$, respectively $W^u_{(c)}$ converges to the linearized map eigenvector $\mathbf{e}^{(e)}$, respectively $\mathbf{e}^{(c)}$. In this sense each eigenvector defines a (curvilinear) axis of the stable, respectively unstable manifold.

Conversely, we can use an arbitrarily small segment of a fixed point eigenvector to construct a finite segment of the associated manifold. Precise construction depends on the type of the eigenvalue(s).

**Expanding real and positive eigendirection**. Consider $i$th expanding eigenvalue, eigenvector pair $(\Lambda_i, \mathbf{e}_i)$ computed from $J$ evaluated at a cycle point,

$$J(x)\mathbf{e}_i(x) = \Lambda_i \mathbf{e}_i(x)\,, \quad x \in p\,, \quad \Lambda_i > 1\,. \tag{11.6}$$

Take an infinitesimal eigenvector $\epsilon\,\mathbf{e}_i(x)$, $\epsilon \ll 1$, and its image $J_p(x)\epsilon\,\mathbf{e}_i(x) = \Lambda_i \epsilon\,\mathbf{e}_i(x)$. Sprinkle the interval $|\Lambda_i - 1|\epsilon$ with a large number of points $x_m$, equidistantly spaced on logarithmic scale $\ln|\Lambda_i - 1| + \ln\epsilon$. The successive images of these points $f(x_j)$, $f^2(x_j)$, $\cdots$, $f^m(x_j)$ trace out the curvilinear unstable manifold in direction $\mathbf{e}_i$. Repeat for $-\epsilon\,\mathbf{e}_i(x)$.

**Contracting real, positive eigendirection**. Reverse the action of the map backwards in time. This turns a contracting direction into an expanding one, tracing out the curvilinear stable manifold in continuation of $\epsilon\,\mathbf{e}_j$.

**Expanding/contracting real negative eigendirection**. As above, but every even iterate $f^2(x_j)$, $f^4(x_j)$, $f^6(x_j)$ continues in the direction $\mathbf{e}_i$, every odd one in the direction $-\mathbf{e}_i$.

**Complex eigenvalue pair**. Construct an orthonormal pair of eigenvectors spanning the plane $\{\epsilon\,\mathbf{e}_j, \epsilon\,\mathbf{e}_{j+1}\}$. Iteration of the annulus between an infinitesimal circle and its image by $J$ spans the spiralling/circle unstable manifold of the complex eigenvalue pair $\{\Lambda_i, \Lambda_{i+1} = \Lambda_i^*\}$.

## 11.3 Horseshoes

If a flow is locally unstable but globally bounded, any open ball of initial points will be stretched out and then folded back. An example is a 3-dimensional invertible flow sketched in figure 10.5 which returns an area of a Poincaré section of the flow stretched and folded into a "horseshoe," such that the initial area is intersected at most twice (see exercise 11.4, the first Figure). Run backwards, the flow generates the backward horseshoe which intersects the forward horseshoe at most 4 times, and so forth. Such flows exist, and are easily constructed–an example is the Rössler flow, discussed in example 3.4. [exercise 11.1]

Now we shall construct an example of a locally unstable but globally bounded mapping which returns 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 the *once-folding* maps.

As an example is afforded by the 2-dimensional *Hénon map* [exercise 3.5]

$$\begin{aligned} x_{n+1} &= 1 - ax_n^2 + by_n \\ y_{n+1} &= x_n \,. \end{aligned} \tag{11.7}$$

The Hénon map models qualitatively the Poincaré section return map of figure 10.5. For $b = 0$ the Hénon map reduces to the parabola (10.7), and, as shown in sects. 3.3 and 27.1, for $b \neq 0$ it is kind of a fattened parabola; by construction, it takes a rectangular initial area and returns it bent as a horseshoe.

For definitiveness, fix the parameter values to $a = 6$, $b = -1$. The map is quadratic, so it has 2 fixed points $x_0 = f(x_0)$, $x_1 = f(x_1)$ indicated in figure 11.3 (a). For the parameter values at hand, they are both unstable. If you start with a small ball of initial points centered around $x_1$, and iterate the map, the ball will be stretched and squashed along the line $W_1^u$. Similarly, a small ball of initial points centered around the other fixed point $x_0$ iterated backward in time,

$$\begin{aligned} x_{n-1} &= y_n \\ y_{n-1} &= -\frac{1}{b}(1 - ay_n^2 - x_n)\,, \end{aligned} \tag{11.8}$$

traces out the line $W_0^s$. $W_0^s$ is the stable manifold of $x_0$ fixed point, and $W_1^u$ is the unstable manifold of $x_1$ fixed point, defined in sect. 11.2. Their intersections enclose the crosshatched region $\mathcal{M}_{\cdot}$ . Any point outside $W_1^u$ border of $\mathcal{M}_{\cdot}$ escapes to infinity forward in time, while any point outside $W_0^s$ border escapes to infinity backwards in time. In this way the unstable - stable manifolds define topologically, invariant and optimal $\mathcal{M}_{\cdot}$ initial region; all orbits that stay confined for all times are confined to $\mathcal{M}_{\cdot}$ .

Iterated one step forward, the region $\mathcal{M}_{\cdot}$ is stretched and folded into a *smale* horseshoe drawn in figure 11.3 (b). The horseshoe fattened parabola shape is the consequence of the quadratic form $x^2$ in (11.7). Parameter $a$ controls the amount of stretching, while the parameter $b$ controls the amount of compression of the folded horseshoe. The case $a = 6$, $b = 0.9$ considered here corresponds to strong stretching and weak compression. Label the two forward intersections $f(\mathcal{M}_{\cdot})\cap\mathcal{M}_{\cdot}$ by $\mathcal{M}_{s\cdot}$, with $s \in \{0, 1\}$, figure 11.3 (b). The horseshoe consists of the two strips $\mathcal{M}_{0\cdot}, \mathcal{M}_{1\cdot}$ , and the bent segment that lies entirely outside the $W_1^u$ line. As 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 $\mathcal{M}_{\cdot}$ is again stretched and folded into a horseshoe, figure 11.3 (c). As stability and instability are interchanged under

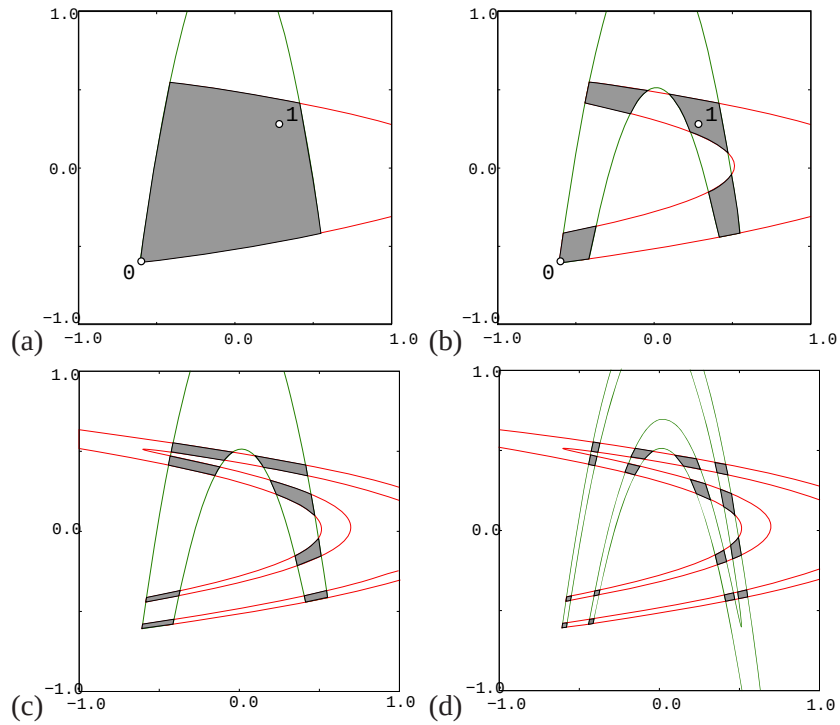



**Figure 11.3:** The Hénon map for $a = 6$, $b = -1$: fixed points $\overline{0}$, $\overline{1}$, with segments of the $W_0^s$ stable manifold, $W_0^u$ unstable manifold. (a) Their intersection bounds the region $\mathcal{M}_{.}$ which contains the non–wandering set $\Omega$. (b) The intersection of the forward image $f^1(\mathcal{M}_{.})$ with the backward backward $f^{-1}(\mathcal{M}_{.})$ is a four-region cover of $\Omega$. (c) The intersection of the twice-folded forward horseshoe $f^2(\mathcal{M}_{.})$ with backward horseshoe $f^{-1}(\mathcal{M}_{.})$. (d) The intersection of $f^2(\mathcal{M}_{.})$ with $f^{-2}(\mathcal{M}_{.})$ is a 16-region cover of $\Omega$. Iteration yields the complete Smale horseshoe non–wandering set $\Omega$, i.e., the union of all non-wandering points of $f$, with every forward fold intersecting every backward fold. (Y. Matsuoka)

time reversal, this horseshoe is transverse to the forward one. Again the points in the horseshoe bend wonder off to infinity as $n \to -\infty$, and we are left with the two (backward) 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 3-disk pinball game figure 10.3. Iterating three steps forward we get an 8 strips, and so on *ad infinitum*.

What is the significance of the subscript $_{.011}$ which labels the $\mathcal{M}_{.011}$ backward 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 forward strips all $x \in \mathcal{M}_{s_{-m}\cdots s_{-1}s_0.}$ have the *past itinerary* $S^- = s_{-m}\cdots s_{-1}s_0$ . Which partition we use to present pictorially 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

$$\mathcal{M}_{0.} = f(\mathcal{M}_{.0}), \qquad \mathcal{M}_{1.} = f(\mathcal{M}_{.1}).$$

$\Omega$, the non–wandering set (2.2) 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 $\mathcal{M}$:

$$\Omega = \left\{ x : x \in \lim_{m,n\to\infty} f^m(\mathcal{M}_{.}) \bigcap f^{-n}(\mathcal{M}_{.}) \right\} . \tag{11.9}$$

Two important properties of the Smale horseshoe are that it has a *complete binary symbolic dynamics* and that it is *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 $S(x) = \cdots s_{-2}s_{-1}s_0.s_1s_2\cdots$, where $s_n = s$ if $f^n(x) \in \mathcal{M}_{.s}$, $s \in \{0,1\}$ and $n \in \mathbb{Z}$. For sufficiently separated disks, the 3-disk game of pinball figure 10.3, is another example of a complete Smale horseshoe; in this case the "folding" region of the horseshoe is cut out of the picture by allowing the pinballs that fly between the disks to fall off the table and escape.

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 $f \to f + \delta$ of the flow, for example, for slight displacements of the disks, 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.

## 11.4 Spatial ordering

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 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 11.8]

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 reverse: 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."

### 11.4.1 Symbol square

For a better visualization of 2-dimensional non–wandering sets, fatten the intersection regions until they completely cover a unit square, as in figure 11.4. 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* $[s_{-m+1}\cdots s_0.s_1s_2\cdots s_n]$. 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 the state space, but in

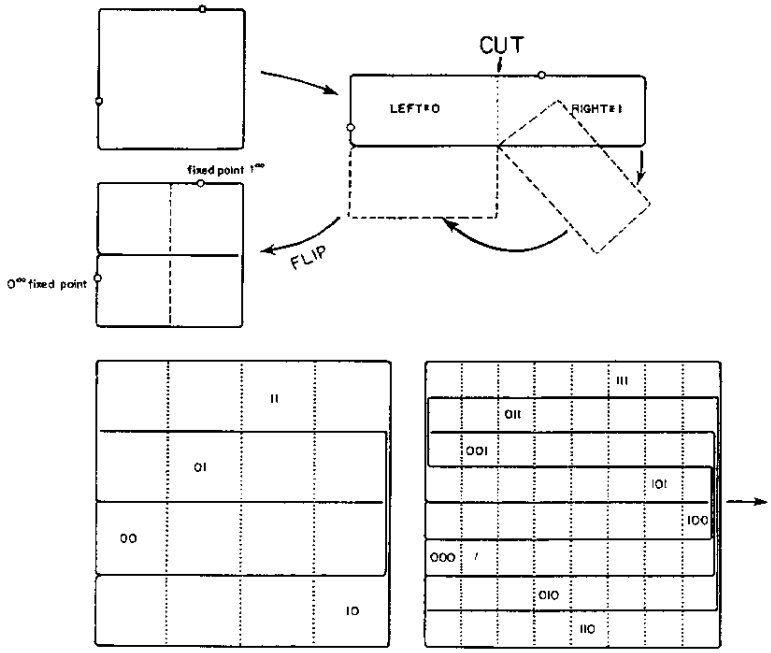



FIG. 4. Iterative construction of the symbol plane.

**Figure 11.4:** Kneading Danish pastry: symbol square representation of an orientation reversing once-folding map obtained by fattening the Smale horseshoe intersections of figure 11.3 into a unit square. In the symbol square the dynamics maps rectangles into rectangles by a decimal point shift.

the symbol square its dynamics is exceedingly simple; all of its points are mapped by the decimal point shift (10.18)

[exercise 11.2]

$$\sigma(\cdots s_{-2}s_{-1}s_0.s_1s_2s_3\cdots) = \cdots s_{-2}s_{-1}s_0s_1.s_2s_3\cdots\,, \tag{11.10}$$

For example, the square [01.01] gets mapped into the rectangle $\sigma[01.01] = [010.1]$, see exercise 11.4, the first Figure (b).

[exercise 11.3]

As the horseshoe mapping is a simple repetitive operation, we expect a simple relation between the symbolic dynamics labeling of the horseshoe strips, and their relative placement. The symbol square points $\gamma(S^+)$ with future itinerary $S^+$ are constructed by converting the sequence of $s_n$'s into a binary number by the algorithm (10.9). This follows by inspection from figure 11.4. In order to understand this relation between the topology of horseshoes and their symbolic dynamics, it might be helpful to backtrace to sect. 10.2.2 and work through and understand first the symbolic dynamics of 1-dimensional unimodal mappings.

[exercise 11.4]

Under backward iteration the roles of 0 and 1 symbols are interchanged; $\mathcal{M}_0^{-1}$ has the same orientation as $\mathcal{M}$, while $\mathcal{M}_1^{-1}$ has the opposite orientation. We assign to an *orientation preserving* once-folding map the *past topological coordinate* $\delta = \delta(S^-)$ by the algorithm:

[exercise 11.5]

$$\begin{aligned} w_{n-1} &= \begin{cases} w_n & \text{if } s_n = 0 \\ 1 - w_n & \text{if } s_n = 1 \end{cases}, \qquad w_0 = s_0 \\ \delta(S^-) &= 0.w_0w_{-1}w_{-2}\ldots = \sum_{n=1}^{\infty} w_{1-n}/2^n\,. \end{aligned} \tag{11.11}$$

Such formulas are best derived by quiet contemplation of the action of a folding map, in the same way we derived the future topological coordinate (10.9).

The coordinate pair $(\delta, \gamma)$ maps a point $(x, y)$ in the state space Cantor set of figure 11.3 into a point in the symbol square of figure 11.4, preserving the topological ordering; $(\delta, \gamma)$ serves as a topologically faithful representation of the non–wandering set of any once-folding map, and aids us in partitioning the set and ordering the partitions for any flow of this type.

## 11.5 Pruning

> 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 méchanique céleste*

In general, not all possible itineraries 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 potential ridge. To count correctly, we need to *prune* the inadmissible trajectories, i.e., specify the grammar of the admissible itineraries.

While the complete Smale horseshoe dynamics discussed so far is rather straightforward, we had to get through it in order to be able to approach a situation that resembles more the real life: 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 13.2 (a). The utility of the symbol square lies in the fact that the surviving, admissible itineraries still maintain the same relative spatial ordering as for the complete case.

In the example of figure 13.2 (a) the rectangles [10.1], [11.1] have been pruned, and consequently *any* trajectory containing blocks $b_1 = 101$, $b_2 = 111$ is pruned. We refer to the border of this primary pruned region as the *pruning front*; another example of a pruning front is drawn in figure 13.2 (d). We call it a "front" as it can be visualized as a border between admissible and inadmissible; any trajectory whose periodic point would fall to the right of the front in figure 13.2 is inadmissible, i.e., pruned. The pruning front is a complete description of the symbolic dynamics of once-folding maps.For now we need this only as a concrete illustration of how pruning rules arise.

In the example at hand there are total of two forbidden blocks 101, 111, so the symbol dynamics is a subshift of finite type (10.22). For now we concentrate on this kind of pruning because it is particularly clean and simple. Unfortunately, for a generic dynamical system a subshift of finite type is the exception rather than the rule. 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, due to the lack of structural stability the transport coefficients such as the deterministic diffusion constant of sect. 24.2 are emphatically

*not* smooth functions of the system parameters. This generic lack of structural stability is what makes nonlinear dynamics so hard.

The conceptually simpler finite subshift Smale horseshoes suffice to motivate most of the key concepts that we shall need for time being.

### 11.5.1 Converting pruning blocks into Markov graphs

The complete binary symbolic dynamics is too simple to be illuminating, so we turn next to the simplest example of pruned symbolic dynamics, the finite subshift obtained by prohibition of repeats of one of the symbols, let us say $_00_$. This situation arises, for example, in studies of the circle maps, where this kind of symbolic dynamics describes "golden mean" rotations. Now the admissible itineraries are enumerated by the pruned binary tree of figure 10.13 (a), or the corresponding Markov graph figure 10.13 (b). We recognize this as the Markov graph example of figure 10.14.

[exercise 13.7]

[exercise 13.8]

So we can already see the main ingredients of a general algorithm: (1) Markov graph encodes 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.

Suppose now that, by hook or crook, you have been so lucky fishing for pruning rules that you now know the grammar (10.21) in terms of a finite set of pruning blocks $\mathcal{G} = \{b_1, b_2, \cdots b_k\}$, of lengths $n_{b_m} \leq M$. Our task is to generate all admissible itineraries. What to do?

**A Markov 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.

2. Label all nodes internal to pruning blocks by the itinerary connecting the root point to the internal node. Why? So far we have pruned forbidden branches by looking $n_b$ steps into future for all pruning blocks. into future for pruning block $b = 10010$. However, the blocks with a 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 internal node all remaining branches allowed by the alphabet, and label them. Why? Each one of them 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. Pick one of the free external nodes 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 the 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 external node in question is [1010.]. Three time steps back 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 external node [1010.] should be identified with the internal node [10.].

5. Repeat until all free nodes have been tied back into the internal nodes.

6. 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.

7. The result is a Markov diagram. There is no guarantee that this is the smartest, most compact Markov diagram possible for given pruning (if you have a better algorithm, teach us), but walks around it do generate all admissible itineraries, and nothing else.

**Heavy pruning.**

We complete this training by examples by implementing the pruning of figure13.2 (d). The pruning blocks are

$$[100.10], [10.1], [010.01], [011.01], [11.1], [101.10]. \tag{11.12}$$

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 13.3 (a). These nodes are the potentially dangerous nodes - beginnings of blocks that might end up pruned. Add the side branches to those nodes (figure 13.3 (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: $00 \to 0$; $110 \to 10$; $011 \to 11$; $0101 \to 101$ (pruned); $1000 \to 00 \to 00 \to 0$; $10011 \to 0011 \to 011 \to 11$; $01000 \to 0$.

As in the previous two examples, the trees originating in identified nodes are identical, so the tree is "self-similar." Now connect the side branches to the corresponding nodes, figure 13.3 (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 Markov graph of figure13.3 (d); the admissible bi-infinite symbol sequences are generated as all possible walks along this graph.

## Résumé

Given a partition $\mathcal{A}$ of the state space $\mathcal{M}$, a dynamical system $(\mathcal{M}, f)$ induces topological dynamics $(\Sigma, \sigma)$ on the space $\Sigma$ of all admissible bi–infinite itineraries. The itinerary describes the time evolution of an orbit, while (for 2-$d$ hyperbolic maps) the symbol square describes the spatial ordering of points along the orbit. The rule that everything to one side of the pruning front is forbidden might (in hindsight) seem obvious, but if you have ever tried to work out symbolic dynamics of some "generic" dynamical system, you should be struck by its simplicity: instead of pruning a Cantor set embedded within some larger Cantor set, the pruning front cleanly cuts out a *compact* region in the symbol square and that is all - there are no additional pruning rules.

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 Markov graphs. These matrices are the simplest examples of evolution operators prerequisite to developing a theory of averaging over chaotic flows.

Importance of symbolic dynamics is often grossly unappreciated; as we shall see in chapters 21 and 18, coupled with uniform hyperbolicity, the existence of a finite grammar is the crucial prerequisite for construction of zeta functions with nice analyticity properties.

## Commentary

**Remark 11.1** Stable/unstable manifolds. For pretty pictures of invariant manifolds other than Lorenz, see Abraham and Shaw [26].

**Remark 11.2** 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 [23], but a physicist 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. If you find yourself mystified by Smale's article abstract about "the action (differentiable) of a Lie group $G$ on a manifold $M$," quoted on page 179, rereading chapter 14 might help; for example, the Liouville operators form a Lie group (of symplectic, or canonical transformations) acting on the manifold $(p, q)$.

**Remark 11.3** Kneading theory. The admissible itineraries are studied in refs. [12, 14, 16, 17], as well as many others. We follow here the Milnor-Thurston exposition [13]. They study the topological zeta function for piecewise monotone maps of the interval, and show that for the finite subshift case it can be expressed in terms of a finite dimensional *kneading determinant*. As the kneading determinant is essentially the topological zeta function that we introduce in (13.4), we shall not discuss it here. Baladi and Ruelle have reworked this theory in a series of papers [15, 16, 17] and in ref. [18] replaced it by a power series manipulation. The kneading theory is covered here in P. Dahlqvist's appendix D.1.

**Remark 11.4** Pruning fronts. The notion of a pruning front was introduced in ref. [19], and developed by K.T. Hansen for a number of dynamical systems in his Ph.D. thesis [8] and a series of papers [26]-[30]. Detailed studies of pruning fronts are carried out in refs. [20, 22, 21]; ref. [5] is the most detailed study carried out so far. The rigorous theory of pruning fronts has been developed by Y. Ishii [23, 24] for the Lozi map, and A. de Carvalho [25] in a very general setting.

**Remark 11.5** The unbearable growth of Markov graphs. A construction of finite Markov partitions is described in refs. [10, 11], as well as in the innumerably 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. 19.3.1. If the image of a trial partition region cuts across only a part of another trial region and thus violates the Markov partition condition (10.4), a further refinement of the partition is needed to distinguish distinct trajectories - figure 13.2 is an example of such refinements.

The finite Markov graph construction sketched above is not necessarily the minimal one; for example, the Markov graph of figure 13.3 does not generate only the "fundamental" cycles (see chapter 18), but shadowed cycles as well, such as $t_{00011}$ in (13.17). For methods of reduction to a minimal graph, consult refs. [6, 51, 7]. Furthermore, when one

implements the time reversed dynamics by the same algorithm, one usually gets a graph of 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-*d* dynamics, but is unnatural for 2-*d* maps whose dynamics it treats as 1-dimensional. In practice, generic pruning grows longer and longer, and more plentiful pruning rules. For generic flows the refinements might never stop, and almost always we might have to deal with infinite Markov partitions, such as those that will be discussed in sect. 13.6. Not only do the Markov 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, and it might look very different form 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 the one of ref. [20]. Still, this seems to be the best technology available, unless the reader alerts us to something superior.

## Exercises

11.1. **A Smale horseshoe.** The Hénon map

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 - ax^2 + y \\ bx \end{bmatrix} \tag{11.13}$$

maps the $(x, y)$ plane into itself - it was constructed by Hénon [2] in order to mimic the Poincaré section of once-folding map induced by a flow like the one sketched in figure 10.5. For definitiveness fix the parameters to $a = 6$, $b = -1$.

a) Draw a rectangle in the $(x, y)$ plane such that its $n$th iterate by the Hénon map intersects the rectangle $2^n$ times.

b) Construct the inverse of the (11.13).

c) Iterate the rectangle back in the time; how many intersections are there between the $n$ forward and $m$ backward iterates of the rectangle?

d) Use the above information about the intersections to guess the $(x, y)$ coordinates for the two fixed points, a 2-cycle point, and points on the two distinct 3-cycles from table **??**. The exact cycle points are computed in exercise 12.10.

11.2. **Kneading Danish pastry.** Write down the $(x, y) \to (x, y)$ mapping that implements the baker's map of figure 11.4, 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 (10.6)).

11.3. **Kneading Danish without flipping.** The baker's map of figure 11.4 includes a flip - a map of this type is called an orientation reversing once-folding map. Write down the $(x, y) \to (x, y)$ 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.

11.4. **Fix this manuscript.** Check whether the layers of the baker's map of figure 11.4 are indeed ordered as the branches of the alternating binary tree of figure 10.9. (They might not be - we have not rechecked them). Draw the correct binary trees that order both the future and past itineraries.

For once-folding maps there are four topologically distinct ways of laying out the stretched and folded image of the starting region,

(a) orientation preserving: stretch, fold upward, as in figure **??**.

(b) orientation preserving: stretch, fold downward, as in figure 13.2

(c) orientation reversing: stretch, fold upward, flip, as in figure **??**.

(d) orientation reversing: stretch, fold downward, flip, as in figure 11.4,

with the corresponding four distinct binary-labeled symbol squares. For $n$-fold "stretch & fold" flows the labeling would be $n$ary. The intersection $\mathcal{M}_0$ for the orientation preserving Smale horseshoe, the first Figure (a) above. is oriented the same way as $\mathcal{M}$, while $\mathcal{M}_1$ is oriented opposite to $\mathcal{M}$. Brief contemplation of figure 11.4 indicates that the forward iteration strips are ordered relative to each other as the branches of the alternating binary tree in figure 10.9.

Check the labeling for all four cases.

11.5. **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 $\gamma$ is identical to the $\gamma$ 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 $\delta$ is given by

$$\begin{aligned} w_{n-1} &= \begin{cases} 1 - w_n & \text{if } s_n = 0 \\ w_n & \text{if } s_n = 1 \end{cases}, \qquad w_0 = s_0 \\ \delta(x) &= 0.w_0 w_{-1} w_{-2} \ldots = \sum_{n=1}^{\infty} w_{1-n}/2^n \end{aligned} \tag{11.14}$$

11.6. **Infinite symbolic dynamics.** Let $\sigma$ be a function that returns zero or one for every infinite binary string: $\sigma : \{0, 1\}^{\mathbb{N}} \to \{0, 1\}$. Its value is represented by $\sigma(\epsilon_1, \epsilon_2, \ldots)$ where the $\epsilon_i$ 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

$$\|a\| = \sup_{\{\epsilon_i\}} |a(\epsilon_1, \epsilon_2, \ldots)| < \infty .$$

For these functions

$$\begin{aligned} \mathcal{T} a(\epsilon_1, \epsilon_2, \ldots) &= a(0, \epsilon_1, \epsilon_2, \ldots)\sigma(0, \epsilon_1, \epsilon_2, \ldots) \\ &\quad + a(1, \epsilon_1, \epsilon_2, \ldots)\sigma(1, \epsilon_1, \epsilon_2, \ldots) . \end{aligned}$$

(a) (easy) Consider a finite version $T_n$ of the operator $\mathcal{T}$:

$$T_n a(\epsilon_1, \epsilon_2, \ldots, \epsilon_{1,n}) = a(0, \epsilon_1, \epsilon_2, \ldots, \epsilon_{n-1})\sigma(0, \epsilon_1, \epsilon_2, \ldots, \epsilon_{n-1}) + a(1, \epsilon_1, \epsilon_2, \ldots, \epsilon_{n-1})\sigma(1, \epsilon_1, \epsilon_2, \ldots, \epsilon_{n-1}) .$$

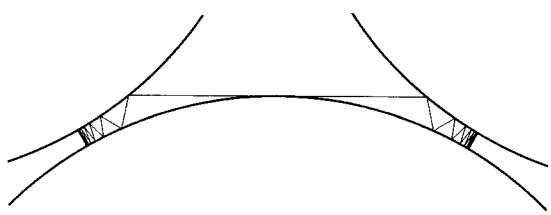
Show that $T_n$ is a $2^n \times 2^n$ matrix. Show that its trace is bounded by a number independent of $n$.

(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.

11.7. **Time reversibility.**** Hamiltonian flows are time reversible. Does that mean that their Markov 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?

11.8. **3-disk pruning** (Not easy) Show that for 3-disk game of pinball the pruning of orbits starts at $R : a = 2.04821419\ldots$.



(Kai T. Hansen)

11.9. **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.

**step 1.** $_1000_$ prunes all cycles with a $_000_$ subsequence with the exception of the fixed point $\overline{0}$; 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. 11.5.1, the alphabet is $\{1, 2, 3; \overline{0}\}$, and the remaining pruning rules have to be rewritten in terms of symbols 2=10, 3=100:

**step 2.** alphabet $\{1, 2, 3; \overline{0}\}$, prune $_33_$, $_213_$, $_313_$. This means that the 3-cycle $\overline{3} = \overline{100}$ is pruned and no long cycles stay close enough to it for a single $_100_$ repeat. As in example 1?!, prohibition of $_33_$ is implemented by dropping the symbol "3" and extending the alphabet by the allowed blocks 13, 23:

**step 3.** alphabet $\{1, 2, \underline{13}, \underline{23}; \overline{0}\}$, prune $_2\underline{13}_$, $_\underline{23}\,\underline{13}_$, $_\underline{13}\,\underline{13}_$, 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 $\{1, 2, \underline{23}, \underline{113}; \overline{0}\}$, unrestricted 4-ary dynamics. The other remaining possible blocks $_\underline{213}_$, $_\underline{2313}_$ are forbidden by the rules of step 3. (continued as exercise 13.20)

## References


[11.1] E. Hopf, *Ergodentheorie* (Chelsea Publ. Co., New York 1948).

[11.2] T. Bedford, M.S. Keane and C. Series, eds., *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces* (Oxford University Press, Oxford, 1991).

[11.3] M.S. Keane, *Ergodic theory and subshifts of finite type*, in ref. [2].

[11.4] B. Kitchens, "Symbolic dynamics, group automorphisms and Markov partition," in *Real and Complex Dynamical Systems*, B. Branner and P. Hjorth, ed. (Kluwer, Dordrecht, 1995).

[11.5] R. Bowen, "Markov partitions for Axiom A diffeomorphisms," *Amer. J. Math.* **92**, 725 (1970).

[11.6] D. Ruelle, *Transactions of the A.M.S.* **185**, 237 (197?).

[11.7] R. Bowen, *Periodic orbits for hyperbolic flows*, Amer. J. Math. **94**, 1-30 (1972).

[11.8] R. Bowen, *Symbolic dynamics for hyperbolic flows*, Amer. J. Math. **95**, 429-460 (1973).

[11.9] R. Bowen and O.E. Lanford, "Zeta functions of restrictions," pp. 43-49 in *Proceeding of the Global Analysis* (A.M.S., Providence 1968).

[11.10] V.M. Alekseev and M.V. Jakobson, *Symbolic dynamics and hyperbolic dynamical systems*, *Phys. Reports* **75**, 287 (1981).

[11.11] A. Manning, "Axiom A diffeomorphisms have rational zeta function," *Bull. London Math. Soc.***3**, 215 (1971).

[11.12] A.N. Sarkovskii, "Coexistence of cycles of a continuous map of a line into itself," *Ukrainian Math. J.* **16**, 61 (1964).

[11.13] J. Milnor and W. Thurston, "On iterated maps of the interval," in A. Dold and B. Eckmann, eds., *Dynamical Systems, Proceedings, U. of Maryland 1986-87, Lec. Notes in Math.* **1342,** 465 (Springer, Berlin 1988).

[11.14] W. Thurston, "On the geometry and dynamics of diffeomorphisms of surfaces," *Bull. Amer. Math. Soc.* **19**, 417 (1988).

[11.15] V. Baladi and D. Ruelle, "An extension of the theorem of Milnor and Thurston on the zeta functions of interval maps," *Ergodic Theory Dynamical Systems* **14**, 621 (1994).

[11.16] V. Baladi, "Infinite kneading matrices and weighted zeta functions of interval maps," *J. Functional Analysis* **128**, 226 (1995).

[11.17] D. Ruelle, "Sharp determinants for smooth interval maps," in F. Ledrappier, J. Lewowicz, and S. Newhouse, eds., *Proceedings of Montevideo Conference 1995* (Addison-Wesley, Harlow 1996).

[11.18] V. Baladi and D. Ruelle, "Sharp determinants," *Invent. Math.* **123**, 553 (1996).

[11.19] P. Cvitanović, G.H. Gunaratne and I. Procaccia, *Phys. Rev.* **A 38**, 1503 (1988).

[11.20] G. D'Alessandro, P. Grassberger, S. Isola and A. Politi, "On the topology of the Hénon Map," *J. Phys.* **A 23**, 5285 (1990).

[11.21] F. Giovannini and A. Politi, "Generating partitions in Hénon-type maps," *Phys. Lett.* **A 161**, 333 (1992);

[11.22] G. D'Alessandro, S. Isola and A. Politi, "Geometric properties of the pruning front," *Prog. Theor. Phys.* **86**, 1149 (1991).

[11.23] Y. Ishii, "Towards the kneading theory for Lozi attractors. I. Critical sets and pruning fronts," Kyoto Univ. Math. Dept. preprint (Feb. 1994).

[11.24] Y. Ishii, "Towards a kneading theory for Lozi mappings. II. A solution of the pruning front conjecture and the first tangency problem," *Nonlinearity* **10**, 731 (1997).

[11.25] A. de Carvalho, Ph.D. thesis, CUNY New York 1995; "Pruning fronts and the formation of horseshoes," preprint (1997).

[11.26] K.T. Hansen, *CHAOS* **2**, 71 (1992).

[11.27] K.T. Hansen, *Nonlinearity* **5**

[11.28] K.T. Hansen, *Nonlinearity* **5**

[11.29] K.T. Hansen, *Symbolic dynamics III, The stadium billiard*, to be submitted to *Nonlinearity*

[11.30] K.T. Hansen, *Symbolic dynamics IV; a unique partition of maps of Hénon type*, in preparation.

[11.31] Fa-Geng Xie and Bai-Lin Hao, "Counting the number of periods in one-dimensional maps with multiple critical points," *Physica A* **202**, 237 (1994).

[11.32] M. Benedicks and L. Carleson, *Ann. of Math.*, **122**, 1 (1985).

[11.33] M. Benedicks and L. Carleson, *IXth Int. Congr. on Mathematical Physics*, B. Simon *et al.*, eds., p.489, (Adam Hilger, Bristol, 1989).

[11.34] M. Benedicks and L. Carleson, *Ann. of Math.* **133**, 73 (1991).

[11.35] G. D'Alessandro and A. Politi, "Hierarchical approach to complexity ...," *Phys. Rev. Lett.* **64**, 1609 (1990).

[11.36] F. Christiansen and A. Politi, "A generating partition for the standard map," *Phys. Rev. E.* **51**, 3811 (1995); arXiv:chao-dyn/9411005

[11.37] F. Christiansen and A. Politi, "Guidelines for the construction of a generating partition in the standard map," *Physica D* **109**, 32 (1997).

[11.38] F. Christiansen and A. Politi, "Symbolic encoding in symplectic maps," *Nonlinearity* **9**, 1623 (1996).

[11.39] F. Christiansen and A. Politi, "Guidelines for the construction of a generating partition in the standard map," *Physica D* **109**, 32 (1997).

[11.40] T. Hall, "Fat one-dimensional representatives of pseudo-Anosov isotopy classes with minimal periodic orbit structure," *Nonlinearity* **7**, 367 (1994).

[11.41] P. Cvitanović and K.T. Hansen, "Symbolic dynamics of the wedge billiard," Niels Bohr Inst. preprint (Nov. 1992)

[11.42] P. Cvitanović and K.T. Hansen, "Bifurcation structures in maps of Hénon type," *Nonlinearity* **11**, 1233 (1998).

[11.43] R.W. Easton, "Trellises formed by stable and unstable manifolds in plane," *Trans. Am. Math. Soc.***294**, 2 (1986).

[11.44] V. Rom-Kedar, "Transport rates of a class of two-dimensional maps and flows," *Physica D* **43**, 229 (1990);

[11.45] V. Daniels, M. Vallières and J-M. Yuan, "Chaotic scattering on a double well: Periodic orbits, symbolic dynamics, and scaling," *Chaos*, **3**, 475, (1993).

[11.46] P.H. Richter, H.-J. Scholz and A. Wittek, "A Breathing Chaos," *Nonlinearity* **1**, 45 (1990).

[11.47] F. Hofbauer, "Periodic points for piecewise monotone transformations," *Ergod. The. and Dynam Sys.* **5**, 237 (1985).

[11.48] F. Hofbauer, "Piecewise invertible dynamical systems," *Prob. Th. Rel. Fields* **72**, 359 (1986).

[11.49] K.T. Hansen, "Pruning of orbits in 4-disk and hyperbola billiards," *CHAOS* **2**, 71 (1992).

[11.50] G. Troll, "A devil's staircase into chaotic scattering," *Pysica D* **50**, 276 (1991)

[11.51] P. Grassberger, "Toward a quantitative theory of self-generated Complexity," *Int. J. Theor. Phys* **25**, 907 (1986).

[11.52] D.L. Rod, *J. Diff. Equ.* **14**, 129 (1973).

[11.53] R.C. Churchill, G. Pecelli and D.L. Rod, *J. Diff. Equ.* **17**, 329 (1975).

[11.54] R.C. Churchill, G. Pecelli and D.L. Rod, in G. Casati and J. Ford, eds., *Como Conf. Proc. on Stochastic Behavior in Classical and Quantum Hamiltonian Systems* (Springer, Berlin 1976).

[11.55] R. Mainieri, Ph. D. thesis, New York University (Aug 1990); *Phys. Rev.* **A 45**,3580 (1992)

[11.56] M.J. Giannoni and D. Ullmo, "Coding chaotic billiards: I. Non-compact billiards on a negative curvature manifold," *Physica* **D 41**, 371 (1990).

[11.57] D. Ullmo and M.J. Giannoni, "Coding chaotic billiards: II. Compact billiards defined on the pseudosphere," *Physica* **D 84**, 329 (1995).

[11.58] H. Solari, M. Natiello and G.B. Mindlin, "Nonlinear Physics and its Mathematical Tools," (IOP Publishing Ltd., Bristol, 1996).

[11.59] R. Gilmore, "Topological analysis of chaotic dynamical systems," submitted to *Rev. Mod. Phys.* (1997).

[11.60] P. Dahlqvist, *On the effect of pruning on the singularity structure of zeta functions*, J. Math. Phys. **38**, 4273 (1997).

[11.61] E. Hille, *Analytic function theory II*, (Ginn and Co., Boston 1962).

# Chapter 12

# Fixed points, and how to get them

HAVING SET UP the dynamical context, now we turn to the key and unavoidable piece of numerics in this subject; search for the solutions $(x, T)$, $x \in \mathbb{R}^d$, $T \in \mathbb{R}$ of the *periodic orbit condition*

$$f^{t+T}(x) = f^t(x)\,, \qquad T > 0 \tag{12.1}$$

for a given flow or mapping.

We know from chapter 16 that cycles are the necessary ingredient for evaluation of spectra of evolution operators. In chapter 10 we have developed a qualitative theory of how these cycles are laid out topologically.

This chapter is intended as a hands-on guide to extraction of periodic orbits, and should be skipped on first reading - you can return to it whenever the need for finding actual cycles arises. 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*. A serious cyclist will want to also learn about the variational methods to find cycles, chapter 27. They are particularly useful when little is understood about the topology of a flow, such as in high-dimensional periodic orbit searches.

[chapter 27]

fast track:
chapter 13, p. 212

A *prime* cycle $p$ of period $T_p$ is a single traversal of the periodic orbit, so our task will be to find a cycle point $x \in p$ and the shortest time $T_p$ for which (12.1) has a solution. A cycle point of a flow $f^t$ which crosses a Poincaré section $n$ times is a fixed point of the $P^n$ iterate of $P$, the return map (3.1), hence we shall refer to all cycles as "fixed points" in this chapter. By cyclic invariance, stability eigenvalues and the period of the cycle are independent of the choice of the initial point, so it will suffice to solve (12.1) at a single cycle point.

[section 5.2]

If the cycle is an attracting limit cycle with a sizable basin of attraction, it can be found by integrating the flow for sufficiently long time. If the cycle is unstable, simple integration forward in time will not reveal it, and 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 a great freedom in constructing such systems, and many different methods are used in practice.

Due to the exponential divergence of nearby trajectories in chaotic dynamical systems, fixed point searches based on direct solution of the fixed-point condition (12.1) as an initial value problem can be numerically very unstable. Methods that start with initial guesses for a number of points along the cycle, such as the multipoint shooting method described here in sect. 12.3, and the variational methods of chapter 27, are considerably more robust and safer. [chapter 27]

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 preparation of a list of all distinct admissible prime periodic symbol sequences, such as the list given in table **??**. The relations between the temporal symbol sequences and the spatial layout of the topologically distinct regions of the state space discussed in chapters 10 and 11 should enable us to guess location of a series of periodic points along a cycle. Armed with such 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?

## 12.1 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 extremize the length of an approximate orbit that visits a desired sequence of boundary bounces.

**Example 12.1** ***Periodic orbits of billiards.*** *Consider how this works for 3-disk pinball game of sect. 11.1. . Label the three disks by* 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 3.2. 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 11.1] [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 $s = (s_1, \ldots, s_n)$, a task that we postpone to sect. 27.3. As an example, the short* [exercise 27.2] [exercise 12.10]

*periods and stabilities of 3-disk cycles computed this way are listed table* **??**, *and some examples are plotted in figure 3.2. It's a no brainer, and millions of such cycles have been computed.*

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. In practice one might be satisfied with any rampaging robot that finds "the most important" cycles. Ergodic exploration of recurrences that we turn to next sometimes perform admirably well.

### 12.1.1 Cycles from long time series

Two wrongs don't make a right, but three lefts do.
—Appliance guru

(L. Rondoni and P. Cvitanović)

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. Ergodic exploration by long-time trajectories (or long-lived transients, in case of strange repellers) can uncover state space regions of low velocity, or finite time recurrences. In addition, such trajectories preferentially sample the natural measure of the 'turbulent' flow, and by initiating searches within the state space concentrations of natural measure bias the search toward the dynamically important invariant solutions. [remark 12.1] [section 14.1]

The search consists of following a long trajectory in state space, and looking for close returns of the trajectory to itself. Whenever the trajectory almost closes in a loop (within a given tolerance), another point of 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 12.1, it is not systematic, it gives no insight into organization of the ergodic sets, and can easily miss very important cycles. Foundations to a systematic exploration of ergodic state space are laid in chapters 10 and 11, but are a bit of work to implement.

### 12.1.2 Cycles found by thinking

Thinking is extra price.
—Argentine saying

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 (2.6) to zero reduces 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 the 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 origin, e.g. the Lorenz $EQ_0 = (0, 0, 0)$ equilibrium.

As to other equilibria: if you have no better idea, create a state space grid, about 50 $x_k$ across $\mathcal{M}$ in each dimension, and compute the velocity field $v_k = v(x_k)$ at each grid point; a few million $v_k$ values are easily stored. Plot $x_k$ for which $|v_k|^2 < \epsilon$, $\epsilon << |v_{max}|^2$ 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 though of as a sequence of near visitations of equilibria. 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 3.2 to maps from a section of the unstable manifold of equilibrium $s_n$ to the section of unstable manifold of equilibrium $s_{n+1}$, and thus reduce 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 covering symbolic dynamics.

invariant

We showed in the Lorenz flow example 10.5 how to reduce the 3-dimensional Lorenz flow to a 1–$d$ return map.

In the Rössler flow example 2.3 we sketched the attractor by running a long chaotic trajectory, and noted that the attractor is very thin, but otherwise the return maps that we plotted were disquieting – figure 3.6 did not appear to be a 1-to-1 map. In the next example we show how to use such information to approximately locate cycles. In the remainder of this chapter and in chapter 27 we shall learn how to turn such guesses into highly accurate cycles.

**Example 12.2** ***Rössler attractor*** *(G. Simon and P. Cvitanović)*

*Run a long simulation of the Rössler flow $f^t$, plot a Poincaré section, as in figure 3.5, and extract the corresponding Poincaré return map $P$, as in figure 3.6. Luck is with us; the figure 12.1 (a) return map $y \to P_1(y, z)$ looks much like a parabola, so we*

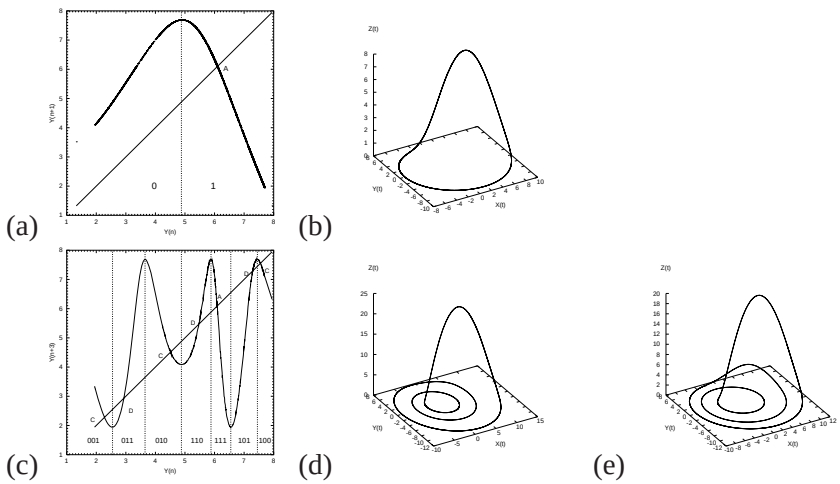



**Figure 12.1:** (a) $y \to P_1(y,z)$ return map for $x = 0, y > 0$ Poincaré section of the Rössler flow figure 2.5. (b) The $\overline{1}$-cycle found by taking the fixed point $y_{k+n} = y_k$ together with the fixed point of the $z \to z$ return map (not shown) an initial guess $(0, y^{(0)}, z^{(0)})$ for the Newton-Raphson search. (c) $y_{k+3} = P_1^3(y_k, z_k)$, the third iterate of Poincaré 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 $\overline{001}$ cycle, and (e) the $\overline{011}$ cycle. (G. Simon)

*take the unimodal map symbolic dynamics, sect. 10.2.1, 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 start resolving its structure.*

*Periodic points of a prime cycle $p$ of cycle length $n_p$ for the $x = 0, y > 0$ Poincaré section of the Rössler flow figure 2.5 are fixed points $(y,z) = P^n(y,z)$ of the $n$th Poincaré return map.*

*Using the fixed point $y_{k+1} = y_k$ in figure 12.1 (a) together with the simultaneous fixed point of the $z \to P_1(y,z)$ return map (not shown) as a starting guess $(0, y^{(0)}, z^{(0)})$ for the Newton-Raphson search for the cycle $p$ with symbolic dynamics label $\overline{1}$, we find the cycle figure 12.1 (b) with the Poincaré section point $(0, y_p, z_p)$, period $T_p$, expanding, marginal, contracting stability eigenvalues $(\Lambda_{p,e}, \Lambda_{p,m}, \Lambda_{p,c})$, and Lyapunov exponents $(\lambda_{p,e}, \lambda_{p,m}, \lambda_{p,c})$:*

[exercise 12.7]

$$
\begin{aligned}
\overline{1}\text{-cycle:} \qquad (x,y,z) &= (0, 6.09176832, 1.2997319) \\
T_1 &= 5.88108845586 \\
(\Lambda_{1,e}, \Lambda_{1,m}, \Lambda_{1,c}) &= (-2.40395353, 1 + 10^{-14}, -1.29 \times 10^{-14}) \\
(\lambda_{1,e}, \lambda_{1,m}, \lambda_{1,c}) &= (0.149141556, 10^{-14}, -5.44)\,. \qquad (12.2)
\end{aligned}
$$

*The Newton-Raphson method that we used is described in sect. 12.4.*

*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 Poincaré return map (3.1) plotted in figure 12.1 (c), together with a corresponding plot for $z_{k+3} = f^3(y_k, z_k)$, to pick starting guesses for the Newton-Raphson searches for the two 3-cycles plotted in figure 12.1 (d), (e). For a listing of the short cycles of the Rössler flow, consult exercise 12.7.*

*The numerical evidence suggests (but 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 the 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.*

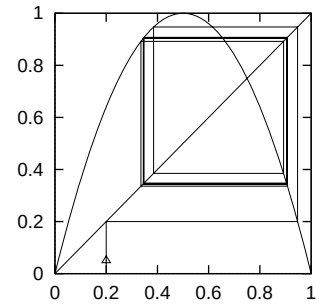



**Figure 12.2:** The inverse time path to the $\overline{01}$-cycle of the logistic map $f(x) = 4x(1-x)$ from an initial guess of $x = 0.2$. At each inverse iteration we chose the 0, respectively 1 branch.

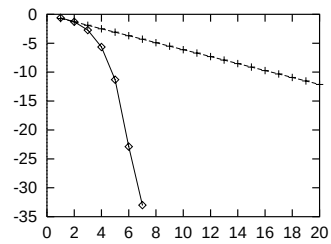



**Figure 12.3:** Convergence of Newton method (◇) vs. inverse iteration (+). The error after $n$ iterations searching for the $\overline{01}$-cycle of the logistic map $f(x) = 4x(1-x)$ with an initial starting guess of $x_1 = 0.2$, $x_2 = 0.8$. $y$-axis is $\log_{10}$ of the error. The difference between the exponential convergence of the inverse iteration method and the super-exponential convergence of Newton method is dramatic.

## 12.2 One-dimensional mappings

(F. Christiansen)

### 12.2.1 Inverse iteration

Let us first consider a very simple method to find unstable cycles of a 1-dimensional map such as the logistic map. Unstable cycles of 1-$d$ 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 branch according to the symbolic dynamics of the cycle we are trying to find, we will automatically converge to the desired cycle. The rate of convergence is given by the stability of the cycle, i.e., the convergence is exponentially fast. Figure 12.2 shows such path to the $\overline{01}$-cycle of the logistic map.

[exercise 12.10]

The method of inverse iteration is fine for finding cycles for 1-d maps and some 2-$d$ systems such as the repeller of exercise 12.10. It is not particularly fast, especially if the inverse map is not known analytically. However, it completely fails for higher dimensional systems where we have both stable and unstable directions. Inverse iteration will exchange these, 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 $f^T(x) = x$.

### 12.2.2 Newton method

Newton method for determining a zero $x^*$ of a function $F(x)$ of one variable is based on a linearization around a starting guess $x_0$:

$$F(x) \approx F(x_0) + F'(x_0)(x - x_0). \tag{12.3}$$

An approximate solution $x_1$ of $F(x) = 0$ is

$$x_1 = x_0 - F(x_0)/F'(x_0). \tag{12.4}$$

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

$$\begin{aligned} x_m &= g(x_{m-1}) = x_{m-1} - F(x_{m-1})/F'(x_{m-1}) \\ &= x_{m-1} - \frac{1}{1 - f'(x_{m-1})}(x_{m-1} - f(x_{m-1}))\,. \end{aligned} \tag{12.5}$$

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 $g$, $g'(x) = 0$, and with a sufficiently good initial guess, the Newton-Raphson iteration will converge super-exponentially fast.

To illustrate the efficiency of the Newton method we compare it to the inverse iteration method in figure 12.3. Newton method wins hands down: the number of significant digits of the accuracy of $x$ estimate doubles with each iteration.

In order to avoid jumping too far from the desired $x^*$ (see figure 12.4), one often initiates the search by the *damped Newton method*,

$$\Delta x_m = x_{m+1} - x_m = -\frac{F(x_m)}{F'(x_m)}\,\Delta\tau\,, \qquad 0 < \Delta\tau \leq 1\,,$$

takes small $\Delta\tau$ steps at the beginning, reinstating to the full $\Delta\tau = 1$ jumps only when sufficiently close to the desired $x^*$.

## 12.3 Multipoint shooting method

(F. Christiansen)

Periodic orbits of length $n$ 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. $f^n$ will be a highly oscillating function with perhaps as many as $2^n$ or more closely spaced fixed points, and finding a specific periodic point, for example 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 length $n$. A better alternative is the *multipoint shooting method*. While it might very hard to give a precise initial point guess for a long periodic orbit, if our guesses are informed by a good state space partition, a rough guess for each point along the desired trajectory might suffice, as for the individual short trajectory segments the errors have no time to explode exponentially.

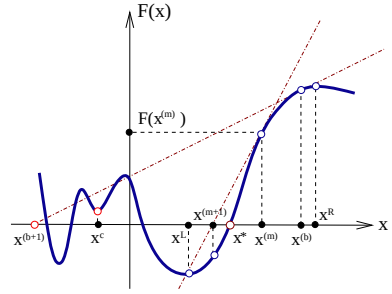



**Figure 12.4:** Newton method: bad initial guess $x^{(b)}$ leads to the Newton estimate $x^{(b+1)}$ far away from the desired zero of $F(x)$. Sequence $\cdots$, $x^{(m)}$, $x^{(m+1)}, \cdots$, starting with a good guess converges super-exponentially to $x^*$. The method diverges if it iterates into the basin of attraction of a local minimum $x^c$.

A cycle of length $n$ is a zero of the $n$-dimensional vector function $F$:

$$F(x) = F\begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 - f(x_n) \\ x_2 - f(x_1) \\ \cdots \\ x_n - f(x_{n-1}) \end{pmatrix}.$$

The relations between the temporal symbol sequences and the spatial layout of the topologically distinct regions of the state space discussed in chapter10 enable us to guess 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 the Newton method now takes the form of

$$\frac{d}{dx}F(x)(x' - x) = -F(x), \tag{12.6}$$

where $\frac{d}{dx}F(x)$ is an $[n \times n]$ matrix:

$$\frac{d}{dx}F(x) = \begin{pmatrix} 1 & & & & -f'(x_n) \\ -f'(x_1) & 1 & & & \\ & \cdots & 1 & & \\ & & \cdots & 1 & \\ & & & -f'(x_{n-1}) & 1 \end{pmatrix}. \tag{12.7}$$

This matrix can easily be inverted numerically by first eliminating the elements below the diagonal. This creates non-zero elements in the $n$th column. We eliminate these and are done.

**Example 12.3** ***Newton inversion for a 3-cycle.*** *Let us illustrate how this works step by step for a 3-cycle. The initial setup for a Newton step is:*

$$\begin{pmatrix} 1 & 0 & -f'(x_3) \\ -f'(x_1) & 1 & 0 \\ 0 & -f'(x_2) & 1 \end{pmatrix}\begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \end{pmatrix} = -\begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix},$$

*where $\Delta x_i = x_i' - x_i$ is the correction to our initial guess $x_i$, and $F_i = x_i - f(x_{i-1})$ is the error at $i$th cycle point. Eliminate the sub-diagonal elements by adding $f'(x_1)$ times the first row to the second row, then adding $f'(x_2)$ times the second row to the third row:*

$$\begin{pmatrix} 1 & 0 & -f'(x_3) \\ 0 & 1 & -f'(x_1)f'(x_3) \\ 0 & 0 & 1 - f'(x_2)f'(x_1)f'(x_3) \end{pmatrix}\begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \end{pmatrix} = -\begin{pmatrix} F_1 \\ F_2 + f'(x_1)F_1 \\ F_3 + f'(x_2)F_2 + f'(x_2)f'(x_1)F_1 \end{pmatrix}.$$

*The next step is to invert the last element in the diagonal, i.e., divide the third row by $1 - f'(x_2)f'(x_1)f'(x_3)$. If this element is zero at the periodic orbit this step cannot work. As $f'(x_2)f'(x_1)f'(x_3)$ is the stability of the cycle (when the Newton iteration has converged), this is not a good method to find marginally stable cycles. We now have*

$$\begin{pmatrix} 1 & 0 & -f'(x_3) \\ 0 & 1 & -f'(x_1)f'(x_3) \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \end{pmatrix} = -\begin{pmatrix} F_1 \\ F_2 + f'(x_1)F_1 \\ \frac{F_3 + f'(x_2)F_2 + f'(x_2)f'(x_1)F_1}{1 - f'(x_2)f'(x_1)f'(x_3)} \end{pmatrix} .$$

*Finally we add $f'(x_3)$ times the third row to the first row and $f'(x_1)f'(x_3)$ times the third row to the second row. The left hand side matrix is now the unit matrix, the right hand side is an explicit formula for the corrections to our initial guess. We have gone through one Newton iteration.*

When one sets up the Newton iteration on the 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, 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 the iteration the vector containing the right hand side should be the correction to the initial guess.

[exercise 12.1]

### 12.3.1 $d$-dimensional mappings

Armed with clever, symbolic dynamics informed initial guesses we can easily extend the Newton-Raphson iteration method to $d$-dimensional mappings. 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 that we went through above we are then manipulating $d$ rows of the left hand side matrix. (Remember that matrices do not commute - always multiply from the left.) In the inversion of the $n$th element of the diagonal we are inverting a $[d \times d]$ matrix $(1 - \prod f'(x_i))$ which can be done if none of the eigenvalues of $\prod f'(x_i)$ equals 1, i.e., if the cycle has no marginally stable eigen-directions.

**Example 12.4** ***Newton method for time delay maps.*** *Some $d$-dimensional mappings (such as the Hénon map (3.18)) can be written as 1-dimensional time delay mappings of the form*

$$f(x_i) = f(x_{i-1}, x_{i-2}, \ldots, x_{i-d}). \tag{12.8}$$

*In this case $\frac{d}{dx}F(x)$ is an $[n \times n]$ matrix as in the case of usual 1-dimensional maps but with non-zero matrix elements on $d$ off-diagonals. In the elimination of these off-diagonal elements the last $d$ columns of the matrix will become non-zero and in the final cleaning of the diagonal we will need to invert a $[d \times d]$ matrix. In this respect, nothing is gained numerically by looking at such maps as 1-dimensional time delay maps.*

## 12.4 Flows

(F. Christiansen)

Further complications arise for flows due to the fact that for a periodic orbit the stability eigenvalue corresponding to the flow direction of necessity equals unity; the separation of any two points along a cycle remains unchanged after a completion of the cycle. More unit eigenvalues can arise if the flow satisfies conservation laws, such as the energy invariance for Hamiltonian systems. We now show how such problems are solved by increasing the number of fixed point conditions. [section 5.2.1]

### 12.4.1 Newton method for flows

A flow is equivalent to a mapping in the sense that one can reduce the flow to a mapping on the Poincaré surface of section. An autonomous flow (2.6) is given as

$$\dot{x} = v(x), \tag{12.9}$$

The corresponding fundamental matrix $M$ (4.43) is obtained by integrating the linearized equation (4.9)

$$\dot{M} = \mathbf{A}M\,, \qquad A_{ij}(x) = \frac{\partial v_i(x)}{\partial x_j}$$

along the trajectory. The flow and the corresponding fundamental matrix are integrated simultaneously, by the same numerical routine. Integrating an initial condition on the Poincaré surface until a later crossing of the same and linearizing around the flow we can write

$$f(x') \approx f(x) + M(x' - x). \tag{12.10}$$

Notice here, that, even though all of $x'$, $x$ and $f(x)$ are on the Poincaré surface, $f(x')$ is usually not. The reason for this is that $M$ corresponds to a specific integration time and has no explicit relation to the arbitrary choice of Poincaré section. This will become important in the extended Newton method described below.

To find a fixed point of the flow near a starting guess $x$ we must solve the linearized equation

$$(1 - M)(x' - x) = -(x - f(x)) = -F(x) \tag{12.11}$$

where $f(x)$ corresponds to integrating from one intersection of the Poincaré surface to another and $M$ is integrated accordingly. Here we run into problems with

the direction along the flow, since - as shown in sect. 5.2.1 - this corresponds to a unit eigenvector of $M$. The matrix $(1-M)$ does therefore not have full rank. A related problem is that the solution $x'$ of (12.11) is not guaranteed to be in the Poincaré surface of section. The two problems are solved simultaneously by adding a small vector along the flow plus an extra equation demanding that $x$ be in the Poincaré surface. Let us for the sake of simplicity assume that the Poincaré surface is a (hyper)-plane, i.e., it is given by the linear equation

$$(x - x_0) \cdot a = 0, \tag{12.12}$$

where $a$ is a vector normal to the Poincaré section and $x_0$ is any point in the Poincaré section. (12.11) then becomes

$$\begin{pmatrix} 1-M & v(x) \\ a & 0 \end{pmatrix} \begin{pmatrix} x'-x \\ \delta T \end{pmatrix} = \begin{pmatrix} -F(x) \\ 0 \end{pmatrix}. \tag{12.13}$$

The last row in this equation ensures that $x$ will be in the surface of section, and the addition of $v(x)\delta T$, a small vector along the direction of the flow, ensures that such an $x$ can be found at least if $x$ is sufficiently close to a solution, i.e., to a fixed point of $f$.

To illustrate this little trick let us take a particularly simple example; consider a 3-d flow with the $(x, y, 0)$-plane as Poincaré section. Let all trajectories cross the Poincaré section perpendicularly, i.e., with $v = (0, 0, v_z)$, 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. In this case the Newton setup looks as follows

$$\begin{pmatrix} 1-\Lambda & 0 & 0 & 0 \\ 0 & 1-\Lambda_s & 0 & 0 \\ 0 & 0 & 0 & v_z \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \delta_x \\ \delta_y \\ \delta_z \\ \delta\tau \end{pmatrix} = \begin{pmatrix} -F_x \\ -F_y \\ -F_z \\ 0 \end{pmatrix}. \tag{12.14}$$

If you consider only the upper-left $[3 \times 3]$ matrix (which is what we would have without the extra constraints that we have introduced) then this matrix is clearly not invertible and the equation does not have a unique solution. However, the full $[4\times4]$ matrix is invertible, as $\det(\cdot) = v_z \det(1-M_\perp)$, where $M_\perp$ is the monodromy matrix for a surface of section transverse to the orbit, see sect. 5.3.

For periodic orbits (12.13) generalizes in the same way as (12.7), but with $n$ additional equations – one for each point on the Poincaré surface. The Newton setup looks like this

$$\left(\begin{array}{ccccc|ccc} 1 & & & & -J_n & v_1 & & \\ -J_1 & 1 & & & & & \ddots & \\ & \cdots & 1 & & & & & v_n \\ & & \cdots & 1 & & & & \\ & & & -J_{n-1} & 1 & & & \\ \hline & a & & & & 0 & & \\ & & \ddots & & & & \ddots & \\ & & & a & & & & 0 \end{array}\right) \begin{pmatrix} \delta_1 \\ \delta_2 \\ \cdot \\ \cdot \\ \delta_n \\ \delta t_1 \\ \cdot \\ \delta t_n \end{pmatrix} = \begin{pmatrix} -F_1 \\ -F_2 \\ \cdot \\ \cdot \\ -F_n \\ 0 \\ \cdot \\ 0 \end{pmatrix}.$$

Solving this equation resembles the corresponding task for maps. However, in the process we will need to invert an $[(d+1)n \times (d+1)n]$ matrix rather than a $[d \times d]$ matrix. The task changes with the length of the cycle.

This method can be extended to take care of the same kind of problems if other eigenvalues of the fundamental matrix equal 1. This happens if the flow has an invariant of motion, the most obvious example being energy conservation in Hamiltonian systems. In this case we add an extra equation for $x$ to be on the energy shell plus and extra variable corresponding to adding a small vector along the gradient of the Hamiltonian. We then have to solve

$$\begin{pmatrix} 1-M & v(x) & \nabla H(x) \\ a & 0 & 0 \end{pmatrix} \begin{pmatrix} x'-x \\ \delta\tau \\ \delta E \end{pmatrix} = \begin{pmatrix} -(x-f(x)) \\ 0 \end{pmatrix} \tag{12.15}$$

simultaneously with

$$H(x') - H(x) = 0. \tag{12.16}$$

The last equation is nonlinear. It is often best to treat this equation separately and solve it in each Newton step. This might mean putting in an additional Newton routine to solve the single step of (12.15) and (12.16) together. One might be tempted to linearize (12.16) and put it into (12.15) to do the two different Newton routines simultaneously, but this will not guarantee a solution on the energy shell. In fact, it may not even be possible to find any solution of the combined linearized equations, if the initial guess is not very good.

### 12.4.2 How good is my orbit?

Provided we understand the topology of the flow, multi-shooting methods and their variational cousins of chapter 27 enable us to compute periodic orbits of arbitrary length. A notion that errors somehow grow exponentially with the cycle length at Lyapunov exponent rate cannot be right. So how do we characterize the accuracy of an orbit of arbitrary length?

The numerical round-off errors along a trajectory are uncorrelated and act as noise, so the errors $(x(t+\Delta t) - f^{\Delta t}(x(t))^2$ are expected to accumulate as the sum of squares of uncorrelated steps, linearly with time. Hence the accumulated numerical noise along an orbit sliced by $N$ intermediate sections separated by $\Delta t_k = t_{k+1} - t_k \sim T_p/N$ can be characterized by an effective diffusion constant

$$D_p = \frac{1}{2(d_e+1)} \sum_{k=1}^{N} \frac{1}{\Delta t_k} (x_{k+1} - f^{\Delta t_k}(x_k))^2 \,. \tag{12.17}$$

For hyperbolic flows errors are exponentially amplified along unstable and contracted along stable eigen-directions, so $d_e + 1$ stands for the number of unstable directions of the flow together with the single marginal direction along the flow. An honest calculation requires an honest error estimate. If you are computing a large set of periodic orbits $p$, list $D_p$ along with $T_p$ and other properties of cycles.

## 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

$$f^{t+T}(x) = f^t(x)\,, \qquad T > 0$$

for a general flow or mapping. Due to the exponential divergence of nearby trajectories in chaotic dynamical systems, direct solution of the periodic orbit condition can be numerically very unstable.

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. 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 regions most frequently visited. The periodic orbit searches can then be initialized by taking nearly recurring orbit segments and deforming them into a closed orbits. With a sufficiently good initial guess the Newton-Raphson formula [section 14.4.1]

$$\begin{pmatrix} 1-M & v(x) \\ a & 0 \end{pmatrix} \begin{pmatrix} \delta x \\ \delta T \end{pmatrix} = \begin{pmatrix} f(x)-x \\ 0 \end{pmatrix}$$

yields improved estimate $x' = x + \delta x, T' = T + \delta T$. Iteration then yields the period $T$ and the location of a periodic point $x_p$ in the Poincaré surface $(x_p - x_0)\cdot a = 0$, where $a$ is a vector normal to the Poincaré section at $x_0$.

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. 12.3, 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 27. [chapter 27]

## Commentary

**Remark 12.1** Close recurrence searches. For low-dimensional maps of 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, ref. [1] is frequently cited. Such methods have been deployed by many, among them

G. Tanner, L. Rondoni, G. Morris, C.P. Dettmann, and R.L. Davidchack [2, 13, 14, 10] (see also sect. 18.5). Sometimes one can determine most of the admissible itineraries and their weights without working too hard, but method comes with no guarantee.

**Remark 12.2** Piecewise linear maps. The Lozi map (3.20) is linear, and 100,000's of cycles can be easily computed by [2×2] matrix multiplication and inversion.

**Remark 12.3** Newton gone wild. Skowronek and Gora [21] offer an interesting discussion of Newton iterations gone wild while searching for roots of polynomials as simple as $x^2 + 1 = 0$.

## Exercises

12.1. **Cycles of the Ulam map.** Test your cycle-searching routines by computing a bunch of short cycles and their stabilities for the Ulam map

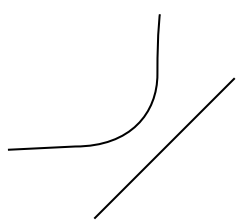
$$f(x) = 4x(1-x)\,. \tag{12.18}$$

12.2. **Cycles stabilities for the Ulam map, exact.** In exercise 12.1 you should have observed that the numerical results for the cycle stability eigenvalues (4.50) are exceptionally simple: the stability eigenvalue 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 (10.6). This problem is perhaps too hard, but give it a try - the answer is in many introductory books on nonlinear dynamics.)

12.3. **Stability of billiard cycles.** Compute stabilities of 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 $L+2$. 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 stabilities for a billiard ball bouncing between the diagonal $y = x$ and one of the hyperbola branches $y = -1/x$.



12.4. **Cycle stability.** Add to the pinball simulator of exercise 8.1 a routine that evaluates the expanding eigenvalue for a given cycle.

12.5. **Pinball cycles.** Determine the stability and length of all fundamental domain prime cycles of the binary symbol string lengths up to 5 (*or longer*) for $R : a = 6$ 3-disk pinball.

12.6. **Newton-Raphson method.** Implement the Newton-Raphson method in 2-*d* and apply it to determination of pinball cycles.

12.7. **Rössler flow cycles.** (continuation of exercise 4.4) Determine all cycles up to 5 Poincaré sections returns for the Rössler flow (2.17), as well as their stabilities.

(Hint: implement (12.13), the multipoint shooting methods for flows; you can cross-check your shortest cycles against the ones listed in the table.)

**Table:** *The Rössler flow (2.17): The itinerary $p$, a periodic point $x_p = (0, y_p, z_p)$ and the expanding eigenvalue $\Lambda_p$ for all cycles up to the topological length 7. (J. Mathiesen, G. Simon, A. Basu)*

| $n_p$ | $p$ | $y_p$ | $z_p$ | $\Lambda_e$ |
|---|---|---|---|---|
| 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 |

12.8. **Cycle stability, helium.** Add to the helium integrator of exercise 2.10 a routine that evaluates the expanding eigenvalue for a given cycle.

12.9. **Colinear helium cycles.** Determine the stability and length of all fundamental domain prime cycles up to symbol sequence length 5 or longer for collinear helium of figure 7.2.

12.10. **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 Poincaré section mappings; can you show that there is exponential contraction to a unique periodic point with a given itinerary? Exercise 27.1 might be helpful in this effort.

12.11. **Inverse iteration method for a Hamiltonian repeller.**

**Table:** *All periodic orbits up to 6 bounces for the Hamiltonian Hénon mapping (12.19) with $a = 6$. Listed are the cycle itinerary, its expanding eigenvalue $\Lambda_p$, and its "center of mass." The "center of mass" is listed because it turns out the "center of mass" is often a simple rational or a quadratic irrational.*

| $p$ | $\Lambda_p$ | $\sum x_{p,i}$ |
|---|---|---|
| 0 | $0.715168\times10^{1}$ | -0.607625 |
| 1 | $-0.295285\times10^{1}$ | 0.274292 |
| 10 | $-0.989898\times10^{1}$ | 0.333333 |
| 100 | $-0.131907\times10^{3}$ | -0.206011 |
| 110 | $0.558970\times10^{2}$ | 0.539345 |
| 1000 | $-0.104430\times10^{4}$ | -0.816497 |
| 1100 | $0.577998\times10^{4}$ | 0.000000 |
| 1110 | $-0.103688\times10^{3}$ | 0.816497 |
| 10000 | $-0.760653\times10^{4}$ | -1.426032 |
| 11000 | $0.444552\times10^{4}$ | -0.606654 |
| 10100 | $0.770202\times10^{3}$ | 0.151375 |
| 11100 | $-0.710688\times10^{3}$ | 0.248463 |
| 11010 | $-0.589499\times10^{3}$ | 0.870695 |
| 11110 | $0.390994\times10^{3}$ | 1.095485 |
| 100000 | $-0.545745\times10^{5}$ | -2.034134 |
| 110000 | $0.322221\times10^{5}$ | -1.215250 |
| 101000 | $0.513762\times10^{4}$ | -0.450662 |
| 111000 | $-0.478461\times10^{4}$ | -0.366025 |
| 110100 | $-0.639400\times10^{4}$ | 0.333333 |
| 101100 | $-0.639400\times10^{4}$ | 0.333333 |
| 111100 | $0.390194\times10^{4}$ | 0.548583 |
| 111010 | $0.109491\times10^{4}$ | 1.151463 |
| 111110 | $-0.104338\times10^{4}$ | 1.366025 |

Consider the Hénon map (3.18) for area-preserving ("Hamiltonian") parameter value $b = -1$. The coordinates of a periodic orbit of length $n_p$ satisfy the equation

$$x_{p,i+1} + x_{p,i-1} = 1 - a x_{p,i}^2 , \quad i = 1, ..., n_p , \tag{12.19}$$

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 (12.19) 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 (12.19)

$$x''_{p,i} = S_{p,i} \sqrt{\frac{1 - x'_{p,i+1} - x'_{p,i-1}}{a}}, \quad i = 1, ..., n_p .$$

Here $S_{p,i}$ are the signs of the corresponding cycle point coordinates, $S_{p,i} = x_{p,i}/|x_{p,i}|$. (G. Vattay)

12.12. **"Center of mass" puzzle**$^{**}$**.** Why is the "center of mass," tabulated in exercise 12.11,

## References


[12.1] D. Auerbach, P. Cvitanović, J.-P. Eckmann, G.H. Gunaratne and I. Procaccia, *Phys. Rev. Lett.* **58**, 2387 (1987).

[12.2] M. Baranger and K.T.R. Davies *Ann. Physics* **177**, 330 (1987).

[12.3] B.D. Mestel and I. Percival, *Physica D* **24**, 172 (1987); Q. Chen, J.D. Meiss and I. Percival, *Physica D* **29**, 143 (1987).

[12.4] find Helleman et all Fourier series methods

[12.5] J.M. Greene, *J. Math. Phys.* **20**, 1183 (1979)

[12.6] H.E. Nusse and J. Yorke, "A procedure for finding numerical trajectories on chaotic saddles" *Physica* **D 36**, 137 (1989).

[12.7] D.P. Lathrop and E.J. Kostelich, "Characterization of an experimental strange attractor by periodic orbits"

[12.8] T. E. Huston, K.T.R. Davies and M. Baranger *Chaos* **2**, 215 (1991).

[12.9] M. Brack, R. K. Bhaduri, J. Law and M. V. N. Murthy, *Phys. Rev. Lett.* **70**, 568 (1993).

[12.10] J.J. Crofts and R.L. Davidchack, "Efficient detection of periodic orbits in chaotic systems by stabilising transformations," arXiv:nlin.CD/0502013.

[12.11] B. Doyon and L. J. Dubé, "On Jacobian matrices for flows," *CHAOS* **15**, 013108 (2005).

[12.12] S.C. Farantos, "Exploring Molecular Vibrational Motions with Periodic Orbits," *Int. Rev. Phys. Chem.* **15**, 345 (1996); tccc.iesl.forth.gr/~farantos, tccc.iesl.forth.gr/articles/review/review1.ps.gz.

[12.13] S.C. Farantos, "POMULT: A Program for Computing Periodic Orbits in Hamiltonian Systems Based on Multiple Shooting Algorithms," *Computer Phys. Comm.* **108**, 240 (1998); esperia.iesl.forth.gr/~farantos/articles/po_cpc/po_ccp.ps.

[12.14] M. Baranger, K.T.R. Davies and J.H. Mahoney, "The calculation of periodic trajectories," *Ann. Phys.* **186**, 95 (1988).

[12.15] K.T.R. Davies, T.E. Huston and M. Baranger, "Calculations of periodic trajectories for the Henon-Heiles Hamiltonian using the monodromy method," *CHAOS* **2**, 215 (1992).

[12.16] N.S. Simonović, "Calculations of periodic orbits: The monodromy method and application to regularized systems," *CHAOS* **9**, 854 (1999).

[12.17] N.S. Simonović, "Calculations of Periodic Orbits for Hamiltonian Systems with Regularizable Singularities," *Few-Body-Systems* **32**, 183 (2003).

[12.18] Z. Gills, C. Iwata, R. Roy, I.B. Scwartz and I. Triandaf, "Tracking Unstable Steady States: Extending the Stability Regime of a Multimode Laser System," *Phys. Rev. Lett.* **69**, 3169 (1992).

[12.19] N.J. Balmforth, P. Cvitanović, G.R. Ierley, E.A. Spiegel and G. Vattay, "Advection of vector fields by chaotic flows," *Stochastic Processes in Astrophysics*, *Annals of New York Academy of Sciences* **706**, 148 (1993); preprint.

[12.20] A. Endler and J.A.C. Gallas, "Rational reductions of sums of orbital coordintes for a Hamiltonian repeller," (2005).

[12.21] L. Skowronek and P. F. Gora, "Chaos in Newtonian iterations: Searching for zeros which are not there," Acta Phys. Polonica B **38**, 1909 (2007); arXiv:nlin/0703061.

# Chapter 13

# Counting

> That which is crooked cannot be made straight: and that which is wanting cannot be numbered.
>
> —Ecclestiastes 1.15

We are now in a position to apply the periodic orbit theory to the first and the easiest problem in theory of chaotic systems: cycle counting. This is the simplest illustration of the raison d'etre of periodic orbit theory; we shall develop a duality transformation that relates *local* information - in this case the next admissible symbol in a symbol sequence - to *global* averages, in this case the mean rate of growth of the number of admissible itineraries with increasing itinerary length. We shall transform the topological dynamics of chapter 10 into a multiplicative operation by means of transition matrices/Markov graphs, and 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 Markov graphs this determinant is a finite polynomial which can be read off the Markov graph.

The method goes well beyond the problem at hand, and forms the core of the entire treatise, making tangible a rather abstract notion of "spectral determinants" yet to come.

## 13.1 How many ways to get there from here?

In the 3-disk system 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 some part 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 $K_n \geq Ce^{n\hat{h}}$. Bounded exponentially by

$3e^{n\ln 2} \geq K_n \geq Ce^{n\hat{h}}$, the number of trajectories must grow exponentially as a function of the itinerary length, with rate given by the *topological entropy*:

$$h = \lim_{n\to\infty} \frac{1}{n} \ln K_n \ . \tag{13.1}$$

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 (10.2) indicates whether the transition from the starting partition $j$ into partition $i$ in one step is allowed or not, and the $(i, j)$ element of the transition matrix iterated $n$ times

[exercise 13.1]

$$(T^n)_{ij} = \sum_{k_1,k_2,\ldots,k_{n-1}} T_{ik_1} T_{k_1k_2} \ldots T_{k_{n-1}j}$$

receives a contribution 1 from every admissible sequence of transitions, so $(T^n)_{ij}$ is the number of admissible $n$ symbol itineraries starting with $j$ and ending with $i$.

**Example 13.1** ***3-disk itinerary counting.***

*The $(T^2)_{13} = 1$ element of $T^2$ for the 3-disk transition matrix (10.5)*

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} . \tag{13.2}$$

*corresponds to $3 \to 2 \to 1$, the only 2-step path from 3 to 1, while $(T^2)_{33} = 2$ counts the two itineraries 313 and 323.*

The total number of admissible itineraries of $n$ symbols is

$$K_n = \sum_{ij} (T^n)_{ij} = (\,1, 1, \ldots, 1\,)\ T^n \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} . \tag{13.3}$$

We can also count the number of prime cycles and pruned periodic points, but in order not to break up the flow of the main argument, we relegate these pretty results to sects. 13.5.2 and 13.7. Recommended reading if you ever have to compute lots of cycles.

The matrix $T$ has non-negative integer entries. A matrix $M$ is said to be *Perron-Frobenius* if some power $k$ of $M$ has strictly positive entries, $(M^k)_{rs} > 0$. In the case of the transition matrix $T$ this means that every partition eventually reaches all of the partitions, i.e., the partition is dynamically transitive or indecomposable, as assumed in (2.2). The notion of *transitivity* is crucial in ergodic theory: a mapping is transitive if it has a dense orbit. This notion is inherited by

the shift operation once we introduce a symbolic dynamics. If that is not the case, state space decomposes into disconnected pieces, each of which can be analyzed separately by a separate indecomposable Markov graph. Hence it suffices to restrict our considerations to transition matrices of Perron-Frobenius type.

A finite $[N\times N]$ matrix $T$ has eigenvalues $T\varphi_\alpha = \lambda_\alpha\varphi_\alpha$ and (right) eigenvectors $\{\varphi_0, \varphi_1, \cdots, \varphi_{M-1}\}$. Expressing the initial vector in (13.3) in this basis (which might be incomplete, $M \le N$),

$$T^n \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = T^n \sum_{\alpha=0}^{N-1} b_\alpha \varphi_\alpha = \sum_{\alpha=0}^{N-1} b_\alpha \lambda_\alpha^n \varphi_\alpha \,,$$

and contracting with $(\,1, 1, \ldots, 1\,)$, we obtain

$$K_n = \sum_{\alpha=0}^{N-1} c_\alpha \lambda_\alpha^n \,.$$

[exercise 13.2]

The constants $c_\alpha$ depend on the choice of initial and final partitions: In this example we are sandwiching $T^n$ between the vector $(\,1, 1, \ldots, 1\,)$ and its transpose, but any other pair of vectors would do, as long as they are not orthogonal to the leading eigenvector $\varphi_0$. In a experiment the vector $(\,1, 1, \ldots, 1\,)$ would be replaced by a description of the initial state,and the right vector would describe the measure time $n$ later.

Perron theorem states that a Perron-Frobenius matrix has a nondegenerate 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, $\lambda_0 > |\mathrm{Re}\,\lambda_\alpha|$, $\alpha = 1, 2, \cdots, N-1$, and the topological entropy (13.1) is given by

$$\begin{aligned} h &= \lim_{n\to\infty} \frac{1}{n} \ln c_0 \lambda_0^n \left[ 1 + \frac{c_1}{c_0} \left( \frac{\lambda_1}{\lambda_0} \right)^n + \cdots \right] \\ &= \ln \lambda_0 + \lim_{n\to\infty} \left[ \frac{\ln c_0}{n} + \frac{1}{n} \frac{c_1}{c_0} \left( \frac{\lambda_1}{\lambda_0} \right)^n + \cdots \right] \\ &= \ln \lambda_0 \,. \end{aligned} \tag{13.4}$$

What have we learned? The transition matrix $T$ is a one-step *short time* operator, advancing the trajectory from a 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 $K_1, K_2, K_3, \ldots$ and estimating (13.1) 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 the free energy is computed from transfer matrix for 1-dimensional lattice models with finite range interactions. Historically, it is analogy with statistical mechanics that led to introduction of evolution operator methods into the theory of chaotic systems.

## 13.2 Topological trace formula

There are two standard ways of getting at eigenvalues of a matrix - by evaluating the trace $\mathrm{tr}\, T^n = \sum \lambda_\alpha^n$, or by evaluating the determinant $\det(1 - zT)$. We start by evaluating the trace of transition matrices.

Consider an $M$-step memory transition matrix, like the 1-step memory example (10.13). The trace of the transition matrix counts the number of partitions that map into themselves. In the binary case the trace picks up only two contributions on the diagonal, $T_{0\cdots0,0\cdots0} + T_{1\cdots1,1\cdots1}$, no matter how much memory we assume. We can even take infinite memory $M \to \infty$, in which case the contributing partitions are shrunk to the fixed points, $\mathrm{tr}\, T = T_{\overline{0},\overline{0}} + T_{\overline{1},\overline{1}}$. [exercise 10.7]

More generally, each closed walk through $n$ concatenated entries of $T$ contributes to $\mathrm{tr}\, T^n$ a product 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 *local trace*, the product of matrix elements along a cycle $c$, each term being multiplied by a book keeping variable $z$. $z^n \mathrm{tr}\, T^n$ is then the sum of $t_c$ for all cycles of length $n$. For example, for an [8×8] transition matrix $T_{s_1 s_2 s_3, s_0 s_1 s_2}$ version of (10.13), 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 $z^3 \mathrm{tr}\, T^3$. This product is manifestly cyclically symmetric, $t_{100} = t_{010} = t_{001}$, and so a prime cycle $p$ of length $n_p$ contributes $n_p$ times, once for each periodic point along its orbit. For the binary labeled non–wandering set the first few traces are given by (consult tables **??** and **??**) [exercise 10.7]

$$
\begin{aligned}
z\,\mathrm{tr}\, T &= t_0 + t_1, \\
z^2 \mathrm{tr}\, T^2 &= t_0^2 + t_1^2 + 2t_{10}, \\
z^3 \mathrm{tr}\, T^3 &= t_0^3 + t_1^3 + 3t_{100} + 3t_{101}, \\
z^4 \mathrm{tr}\, T^4 &= t_0^4 + t_1^4 + 2t_{10}^2 + 4t_{1000} + 4t_{1001} + 4t_{1011}.
\end{aligned} \tag{13.5}
$$

For complete binary symbolic dynamics $t_p = z^{n_p}$ for every binary prime cycle $p$; if there is pruning $t_p = z^{n_p}$ if $p$ is admissible cycle and $t_p = 0$ otherwise. Hence $\mathrm{tr}\, T^n$ counts the number of *admissible periodic points* of period $n$. In general, the $n$th order trace (13.5) picks up contributions from all repeats of prime cycles, with each cycle contributing $n_p$ periodic points, so the total number of periodic points of period $n$ is given by

$$
z^n N_n = z^n \mathrm{tr}\, T^n = \sum_{n_p | n} n_p t_p^{n/n_p} = \sum_p n_p \sum_{r=1}^{\infty} \delta_{n,n_p r} t_p^r \,. \tag{13.6}
$$

Here $m|n$ means that $m$ is a divisor of $n$, and (taking $z = 1$) $t_p = 1$ if the cycle is admissible, and $t_p = 0$ otherwise.

In order to get rid of the awkward divisibility constraint $n = n_p r$ in the above

**Table 13.1:** The total numbers of periodic points $N_n$ of period $n$ for binary symbolic dynamics. The numbers of prime cycles contributing illustrates the preponderance of long prime cycles of length $n$ over the repeats of shorter cycles of lengths $n_p$, $n = rn_p$. Further listings of binary prime cycles are given in tables **??** and **??**. (L. Rondoni)

| $n$ | $N_n$ | # of prime cycles of length $n_p$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 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, we introduce the generating function for numbers of periodic points

$$\sum_{n=1}^{\infty} z^n N_n = \mathrm{tr}\,\frac{zT}{1-zT} \ . \tag{13.7}$$

Substituting (13.6) into the left hand side, and replacing the right hand side by the eigenvalue sum $\mathrm{tr}\,T^n = \sum \lambda_\alpha^n$, we obtain our first example of a trace formula, the *topological trace formula*

$$\sum_{\alpha=0} \frac{z\lambda_\alpha}{1-z\lambda_\alpha} = \sum_p \frac{n_p t_p}{1-t_p} \ . \tag{13.8}$$

A trace formula relates the spectrum of eigenvalues of an operator - in this case the transition matrix - to the spectrum of periodic orbits of the dynamical system. The $z^n$ sum in (13.7) is a discrete version of the Laplace transform (see chapter 16), and the resolvent on the left hand side is the antecedent of the more sophisticated trace formulas (16.10) and (16.23).We shall now use this result to compute the spectral determinant of the transition matrix.

## 13.3 Determinant of a graph

Our next task is to determine the zeros of the *spectral determinant* of an $[M \times M]$ transition matrix

$$\det(1-zT) = \prod_{\alpha=0}^{M-1} (1-z\lambda_\alpha) \ . \tag{13.9}$$

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 18 for arbitrary dynamical averages. For $T$ a finite matrix, (13.9) is just the characteristic equation for $T$. However, we shall be able to compute this object even when the dimension of $T$ and other such operators goes to $\infty$, and for that reason we prefer to refer to (13.9) loosely as the "spectral determinant."

There are various definitions of the determinant of a matrix; they mostly reduce to the statement that the determinant is a certain sum over all possible permutation cycles composed of the traces $\mathrm{tr}\, T^k$, in the spirit of the determinant–trace relation (1.15):

[exercise 4.1]

$$\begin{aligned}\det(1-zT) &= \exp(\mathrm{tr}\ \ln(1-zT)) = \exp\left(-\sum_{n=1}\frac{z^n}{n}\mathrm{tr}\, T^n\right)\\ &= 1-z\,\mathrm{tr}\, T-\frac{z^2}{2}\left((\mathrm{tr}\ \ T)^2-\mathrm{tr}\,(T^2)\right)-\dots \qquad (13.10)\end{aligned}$$

This is sometimes called a cumulant expansion. Formally, the right hand is an infinite sum over powers of $z^n$. 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 $z^n$, $n>M$ must vanish *exactly*.

We now proceed to relate the determinant in (13.10) to the corresponding Markov graph of chapter 10: to this end we start by the usual algebra textbook expression for a determinant as the sum of products of all permutations

$$\det(1-zT) = \sum_{\{\pi\}}(-1)^{\pi}(1-zT)_{1,\pi_1}(1-zT)_{2,\pi_2}\cdots(1-zT)_{M,\pi_M} \qquad (13.11)$$

where $T$ is a $[M\times M]$ matrix, $\{\pi\}$ denotes the set of permutations of $M$ symbols, $\pi_k$ is what $k$ is permuted into by the permutation $\pi$, and $(-1)^{\pi}=\pm1$ is the parity of permutation $\pi$. The right hand side of (13.11) yields a polynomial of order $M$ in $z$: a contribution of order $n$ in $z$ picks up $M-n$ unit factors along the diagonal, the remaining matrix elements yielding

$$(-z)^n(-1)^{\tilde{\pi}}T_{\eta_1,\tilde{\pi}_{\eta_1}}\cdots T_{\eta_n,\tilde{\pi}_{\eta_n}} \qquad (13.12)$$

where $\tilde{\pi}$ is the permutation of the subset of $n$ distinct symbols $\eta_1\dots\eta_n$ indexing $T$ matrix elements. As in (13.5), we refer to any combination $t_c=T_{\eta_1\eta_2}T_{\eta_2\eta_3}\cdots T_{\eta_k\eta_1}$, for a given itinerary $c=\eta_1\eta_2\cdots,\eta_k$, as the *local trace* associated with a closed loop c on the Markov graph. Each term of form (13.12) may be factored in terms of local traces $t_{c_1}t_{c_2}\cdots t_{c_k}$, that is loops on the Markov graph. These loops are non-intersecting, as each node may only be reached by *one* link, and they are indeed loops, as if a node is reached by a link, it has to be the starting point of another *single* link, as each $\eta_j$ must appear exactly *once* as a row and column index.

So the general structure is clear, a little more thinking is only required to get the sign of a generic contribution. We consider only the case of loops of length 1 and 2, and leave to the reader the task of generalizing the result by induction. Consider first a term in which only loops of unit length appear on (13.12), that is, only the diagonal elements of $T$ are picked up. We have $k = n$ loops and an even permutation $\tilde{\pi}$ so the sign is given by $(-1)^k$, $k$ being 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 (13.12) gives $(-1)^n = (-1)^{2j+i}$. So once again these terms combine into $(-1)^k$, where $k = i + j$ is the number of loops. We may summarize our findings as follows:

[exercise 13.3]

> The characteristic polynomial of a transition matrix/Markov graph is given by the sum of all possible partitions $\pi$ of the graph into products of non-intersecting loops, with each loop trace $t_p$ carrying a minus sign:
>
> $$\det(1 - zT) = \sum_{k=0}^{f} \sum_{\pi}{}' (-1)^k t_{p_1} \cdots t_{p_k} \tag{13.13}$$

Any self-intersecting loop is *shadowed* by a product of two loops that share the intersection point. As both the long loop $t_{ab}$ and its shadow $t_a t_b$ in the case at hand carry the same weight $z^{n_a+n_b}$, the cancellation is exact, and the loop expansion (13.13) is finite, with $f$ the maximal number of non-intersecting loops.

We refer to the set of all non-self-intersecting loops $\{t_{p_1}, t_{p_2}, \cdots t_{p_f}\}$ as the *fundamental cycles*. This is not a very good definition, as the Markov graphs are not unique – the most we know is that for a given finite-grammar language, there exist Markov graph(s) with the minimal number of loops. Regardless of how cleverly a Markov graph is constructed, it is always true that for any finite Markov graph the number of fundamental cycles $f$ is finite. If you know a better way to define the "fundamental cycles," let us know.

fast track:
sect. 13.4, p. 220

### 13.3.1 Topological polynomials: learning by examples

The above definition of the determinant in terms of traces is most easily grasped by working through a few examples. The complete binary dynamics Markov graph of figure 10.11 (b) is a little bit too simple, but let us start humbly.

**Example 13.2** ***Topological polynomial for complete binary dynamics:*** *There are only two non-intersecting loops, yielding*

$$\det(1 - zT) = 1 - t_0 - t_1 = 1 - 2z\,. \tag{13.14}$$

*The leading (and only) zero of this characteristic polynomial yields the topological entropy $e^h = 2$. As we know that there are $K_n = 2^n$ binary strings of length $N$, we are not surprised by this result.*

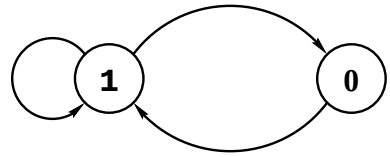



**Figure 13.1:** The golden mean pruning rule Markov graph, see also figure 10.13.

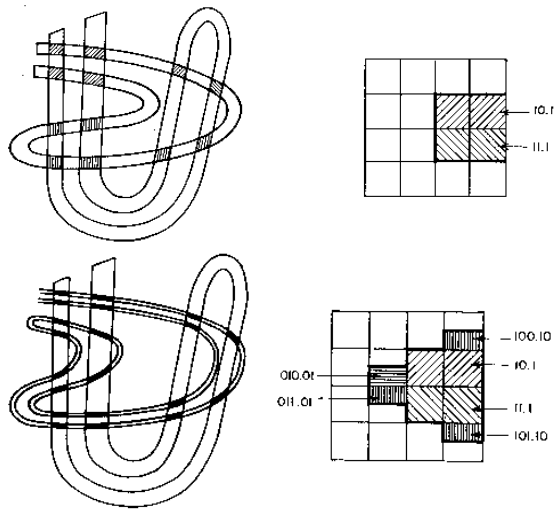


**Figure 13.2:** (a) An incomplete Smale horseshoe: the inner forward fold does not intersect the two rightmost backward folds. (b) The primary pruned region in the symbol square and the corresponding forbidden binary blocks. (c) An incomplete Smale horseshoe which illustrates (d) the monotonicity of 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 figures (a) and (c) is only schematic - in invertible mappings there are further missing intersections, all obtained by the forward and backward iterations of the primary pruned region.

Similarly, for complete symbolic dynamics of $N$ symbols the Markov graph has one node and $N$ links, yielding

$$\det(1 - zT) = 1 - Nz\,, \tag{13.15}$$

whence the topological entropy $h = \ln N$.

**Example 13.3** ***Golden mean pruning:*** *A more interesting example is the "golden mean" pruning of figure 13.1. There is only one grammar rule, that a repeat of symbol 0 is forbidden. The non-intersecting loops are of length 1 and 2, so the topological polynomial is given by* [exercise 13.4]

$$\det(1 - zT) = 1 - t_1 - t_{01} = 1 - z - z^2\,. \tag{13.16}$$

*The leading root of this polynomial is the golden mean, so the entropy (13.4) is the logarithm of the golden mean,* $h = \ln \frac{1+\sqrt{5}}{2}$.

**Example 13.4** ***Nontrivial pruning:*** *The non-self-intersecting loops of the Markov graph of figure 13.3 (d) are indicated in figure 13.3 (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:*

$$\begin{aligned}\det(1 - zT) &= 1 - t_0 - t_{0011} - t_{0001} - t_{00011} \\ &\quad + t_0 t_{0011} + t_{0011} t_{0001}\,.\end{aligned} \tag{13.17}$$

*With* $t_p = z^{n_p}$, *where* $n_p$ *is the length of the* $p$*-cycle, the smallest root of*

$$0 = 1 - z - 2z^4 + z^8 \tag{13.18}$$

*yields the topological entropy* $h = -\ln z$, $z = 0.658779\ldots$, $h = 0.417367\ldots$, *significantly smaller than the entropy of the covering symbolic dynamics, the complete binary shift* $h = \ln 2 = 0.693\ldots$

[exercise 13.9]

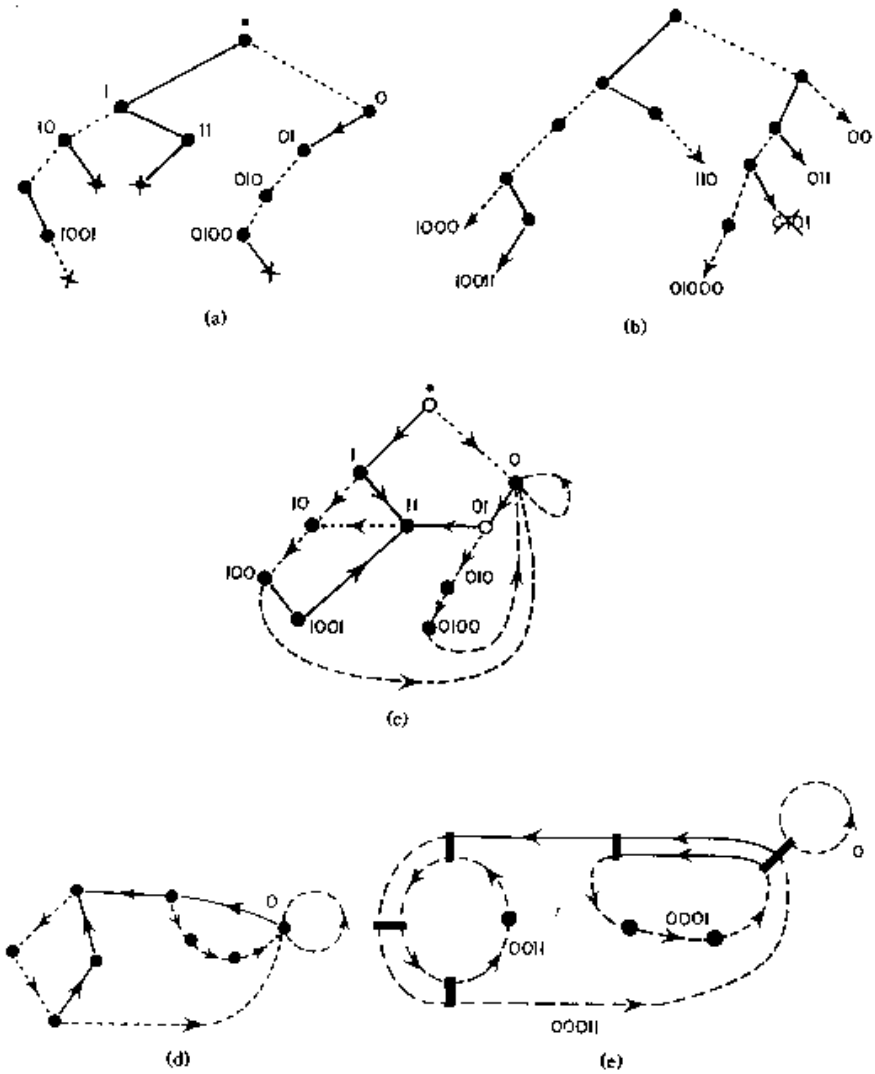



**Figure 13.3:** Conversion of the pruning front of figure 13.2 (d) into a finite Markov graph. (a) Starting with the start node ".", delineate all pruning blocks on the binary tree. A solid line stands for "1" and a dashed line for "0". Ends of forbidden strings are marked with ×. Label all internal nodes by reading the bits connecting ".", the base of the tree, to the node. (b) Indicate all admissible starting blocks by arrows. (c) Drop recursively the leading bits in the admissible blocks; if the truncated string corresponds to an internal node in (a), connect them. (d) Delete the transient, non-circulating nodes; all admissible sequences are generated as walks on this finite Markov graph. (e) Identify all distinct loops and construct the determinant (13.17).

## 13.4 Topological zeta function

What happens if there is no finite-memory transition matrix, if the Markov graph is infinite? If we are never sure that looking further into future will reveal no 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 (13.10)

[exercise 4.1]

$$\det(1 - zT) = 1 - \sum_{n=1}^{\infty} \hat{c}_n z^n \,. \tag{13.19}$$

For finite dimensional matrices the expansion is a finite polynomial, and (13.19) is an identity; 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 (13.9) in

terms of cycles, substitute (13.6) into (13.19) and sum over the repeats of prime cycles using $\ln(1-x) = \sum_r x^r/r$,

$$\det(1-zT) = \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{t_p^r}{r}\right) = \prod_p (1-t_p)\,, \tag{13.20}$$

where for the topological entropy the weight assigned to a prime cycle $p$ of length $n_p$ 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 the *Artin-Mazur* zeta function, conventionally denoted by

$$1/\zeta_{\rm top} = \prod_p (1-z^{n_p}) = 1 - \sum_{n=1} \hat{c}_n z^n \ . \tag{13.21}$$

Counting cycles amounts to giving each admissible prime cycle $p$ weight $t_p = z^{n_p}$ and expanding the Euler product (13.21) as a power series in $z$. As the precise expression for coefficients $\hat{c}_n$ in terms of local traces $t_p$ is more general than the current application to counting, we shall postpone its derivation to chapter18.

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 (13.13) is finite, and we refer to this expansion as the *topological polynomial* of order $\leq M$. The power of defining a determinant by the cumulant expansion is that it works even when the partition is infinite, $M \to \infty$; an example is given in sect. 13.6, and many more later on.

fast track:
sect. 13.6, p. 226

### 13.4.1 Topological zeta function for flows

We now apply the method that we shall use in deriving (16.23) to the problem of deriving the topological zeta functions for flows. The time-weighted density of prime cycles of period $t$ is

$$\Gamma(t) = \sum_p \sum_{r=1} T_p\, \delta(t - rT_p)\,. \tag{13.22}$$

As in (16.22), a Laplace transform smooths the sum over Dirac delta spikes and yields the *topological trace formula*

$$\sum_p \sum_{r=1} T_p \int_{0_+}^{\infty} dt\, e^{-st}\, \delta(t - rT_p) = \sum_p T_p \sum_{r=1}^{\infty} e^{-sT_p r} \tag{13.23}$$

and the *topological zeta function* for flows:

$$1/\zeta_{\rm top}(s) \;=\; \prod_p \left(1 - e^{-sT_p}\right) , \tag{13.24}$$

related to the trace formula by

$$\sum_p T_p \sum_{r=1}^{\infty} e^{-sT_p r} \;=\; -\frac{\partial}{\partial s} \ln 1/\zeta_{\rm top}(s) \,.$$

This is the continuous time version of the discrete time topological zeta function (13.21) for maps; its leading zero $s = -h$ yields the topological entropy for a flow.

## 13.5 Counting cycles

In what follows we shall occasionally need to compute all cycles up to topological length $n$, so it is handy to know their exact number.

### 13.5.1 Counting periodic points

$N_n$, the number of periodic points of period $n$ can be computed from (13.19) and (13.7) as a logarithmic derivative of the topological zeta function

$$\begin{aligned} \sum_{n=1} N_n z^n \;&=\; \mathrm{tr}\left(-z\frac{d}{dz}\ln(1-zT)\right) = -z\frac{d}{dz}\ln\det\,(1-zT) \\ &=\; \frac{-z\frac{d}{dz}1/\zeta_{\rm top}}{1/\zeta_{\rm top}} \,. \end{aligned} \tag{13.25}$$

We see that the trace formula (13.8) diverges at $z \to e^{-h}$, as the denominator has a simple zero there.

**Example 13.5** ***Complete $N$-ary dynamics:*** *As a check of formula (13.19) in the finite grammar context, consider the complete $N$-ary dynamics (10.3) for which the number of periodic points of period $n$ is simply $\mathrm{tr}\, T_c^n = N^n$. Substituting*

$$\sum_{n=1}^{\infty} \frac{z^n}{n} \mathrm{tr}\, T_c^n = \sum_{n=1}^{\infty} \frac{(zN)^n}{n} = \ln(1-zN) \,,$$

*into (13.19) we verify (13.15). The logarithmic derivative formula (13.25) in this case does not buy us much either, we recover*

$$\sum_{n=1} N_n z^n = \frac{Nz}{1-Nz} \,.$$

**Example 13.6** ***Nontrivial pruned dynamics:*** *Consider the pruning of figure 13.3 (e). Substituting (13.18) we obtain*

$$\sum_{n=1} N_n z^n = \frac{z + 8z^4 - 8z^8}{1 - z - 2z^4 + z^8} \,. \tag{13.26}$$

*Now the topological zeta function is not merely a tool for extracting the asymptotic growth of $N_n$; it actually yields the exact and not entirely trivial recursion relation for the numbers of periodic points: $N_1 = N_2 = N_3 = 1$, $N_n = 2n + 1$ for $n = 4, 5, 6, 7, 8$, and $N_n = N_{n-1} + 2N_{n-4} - N_{n-8}$ for $n > 8$.*

### 13.5.2 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 out 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, $p = \overline{011} = \overline{101} = \overline{110} = \ldots 011011 \ldots$ is prime, but $\overline{0101} = 010101 \ldots = \overline{01}$ 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 length $n$ is therefore related to the number of prime cycles by

$$N_n = \sum_{d|n} d M_d \,, \tag{13.27}$$

where $N_n$ equals $\mathrm{tr}\, T^n$. The number of prime cycles can be computed recursively

$$M_n = \frac{1}{n}\left( N_n - \sum_{d|n}^{d<n} d M_d \right) ,$$

or by the *Möbius inversion formula*

[exercise 13.10]

$$M_n = n^{-1} \sum_{d|n} \mu\left(\frac{n}{d}\right) N_d \,. \tag{13.28}$$

where the Möbius function $\mu(1) = 1$, $\mu(n) = 0$ if $n$ has a squared factor, and $\mu(p_1 p_2 \ldots p_k) = (-1)^k$ if all prime factors are different.

We list the number of prime cycles up to length 10 for 2-, 3- and 4-letter complete symbolic dynamics in table **??**. The number of *prime* cycles follows by Möbius inversion (13.28).

[exercise 13.11]

**Table 13.2:** Number of prime cycles for various alphabets and grammars up to length 10. The first column gives the cycle length, the second the formula (13.28) for the number of prime cycles for complete $N$-symbol dynamics, columns three through five give the numbers 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 |

**Example 13.7** ***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 $N_n = \mathrm{tr}\, T^n$. The pruning of self-bounces eliminates the diagonal entries, $T_{N-disk} = T_c - \mathbf{1}$, so the number of the $N$-disk periodic points is*

$$N_n = \mathrm{tr}\, T^n_{N-disk} = (N-1)^n + (-1)^n(N-1) \tag{13.29}$$

*(here $T_c$ is the complete symbolic dynamics transition matrix (10.3)). For the $N$-disk pruned case (13.29) Möbius inversion (13.28) yields*

$$\begin{aligned} M_n^{N-disk} &= \frac{1}{n}\sum_{d|n}\mu\left(\frac{n}{d}\right)(N-1)^d + \frac{N-1}{n}\sum_{d|n}\mu\left(\frac{n}{d}\right)(-1)^d \\ &= M_n^{(N-1)} \quad \text{for} \quad n > 2\ . \end{aligned} \tag{13.30}$$

*There are no fixed points, $M_1^{N-disk} = 0$. The number of periodic points of period 2 is $N^2 - N$, hence there are $M_2^{N-disk} = N(N-1)/2$ prime cycles of length 2; for lengths $n > 2$, the number of prime cycles is the same as for the complete $(N-1)$-ary dynamics of table* **??**.

**Example 13.8** ***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 (eq (13.30)). 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 $(1-z^2t_{12})(1-z^2t_{13})(1-z^2t_{23})$:*

[exercise 13.14]

[exercise 13.15]

$$\begin{aligned} 1/\zeta_{3-disk} &= (1-2z)\frac{(1-z^2)^3}{(1-z)^2(1-z^2)} \\ &= (1-2z)(1+z)^2 = 1-3z^2-2z^3\ . \end{aligned} \tag{13.31}$$

*The factorization reflects the underlying 3-disk symmetry; we shall rederive it in (19.25). As we shall see in chapter 19, symmetries lead to factorizations of topological polynomials and topological zeta functions.*

**Table 13.3:** List of the 3-disk prime cycles up to length 10. Here $n$ is the cycle length, $M_n$ the number of prime cycles, $N_n$ the number of periodic points and $S_n$ the number of distinct prime cycles under the $C_{3v}$ symmetry (see chapter 19 for further details). Column 3 also indicates the splitting of $N_n$ into contributions from orbits of lengths 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 was chosen. The cycles of length 9 grouped by parenthesis are related by time reversal symmetry, but not by any other $C_{3v}$ 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 | |

**Table 13.4:** List of the 4-disk prime cycles up to length 8. The meaning of the symbols is the same as in table **??**. Orbits related by time reversal symmetry (but no other symmetry) already appear at cycle length 5. List of the cycles of length 7 and 8 has been omitted.

| n | $M_n$ | $N_n$ | $S_n$ | $m_p \cdot \hat{p}$ |
|---|---|---|---|---|
| 1 | 0 | 0 | 0 | |
| 2 | 6 | 12=6·2 | 2 | 4·12 + 2·13 |
| 3 | 8 | 24=8·3 | 1 | 8·123 |
| 4 | 18 | 84=6·2+18·4 | 4 | 8·1213 + 4·1214 + 2·1234 + 4·1243 |
| 5 | 48 | 240=48·5 | 6 | 8·(12123 + 12124) + 8·12313 |
| | | | | + 8·(12134 + 12143) + 8·12413 |
| 6 | 116 | 732=6·2+8·3+116·6 | 17 | 8·121213 + 8·121214 + 8·121234 |
| | | | | + 8·121243 + 8·121313 + 8·121314 |
| | | | | + 4·121323 + 8·(121324 + 121423) |
| | | | | + 4·121343 + 8·121424 + 4·121434 |
| | | | | + 8·123124 + 8·123134 + 4·123143 |
| | | | | + 4·124213 + 8·124243 |
| 7 | 312 | 2184 | 39 | |
| 8 | 810 | 6564 | 108 | |

**Example 13.9** ***Alphabet*** $\{a, cb^k;\ \overline{b}\}$***:*** *(continuation of exercise 13.16) In the cycle counting case, the dynamics in terms of* $a \to z$, $cb^k \to \frac{z}{1-z}$ *is a complete binary dynamics with the explicit fixed point factor* $(1 - t_b) = (1 - z)$*:* [exercise 13.16]

$$1/\zeta_{top} = (1-z)\left(1 - z - \frac{z}{1-z}\right) = 1 - 3z + z^2 .$$

[exercise 13.19]

## 13.6 Topological zeta function for an infinite partition

(K.T. Hansen and P. Cvitanović)

Now consider an example of a dynamical system which (as far as we know - there is no proof) has an infinite partition, or an infinity of longer and longer pruning rules. Take the 1-$d$ quadratic map

$$f(x) = Ax(1-x)$$

with $A = 3.8$. It is easy to check numerically that the itinerary or the "kneading sequence" of the critical point $x = 1/2$ is

$$K = 10110111101101111010111101111110\ldots$$

where the symbolic dynamics is defined by the partition of figure 10.6. 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, to give you a feeling for what a "typical" infinite partition topological zeta function looks like. Approximating the dynamics by a Markov graph corresponding to a repeller of the period 29 attractive cycle close to the $A = 3.8$ strange attractor yields a Markov graph with 29 nodes and the characteristic polynomial

$$\begin{aligned} 1/\zeta^{(29)}_{\text{top}} &= 1 - z^1 - z^2 + z^3 - z^4 - z^5 + z^6 - z^7 + z^8 - z^9 - z^{10} \\ &\quad + z^{11} - z^{12} - z^{13} + z^{14} - z^{15} + z^{16} - z^{17} - z^{18} + z^{19} + z^{20} \\ &\quad - z^{21} + z^{22} - z^{23} + z^{24} + z^{25} - z^{26} + z^{27} - z^{28} . \end{aligned} \tag{13.32}$$

The smallest real root of this approximate topological zeta function is [exercise 13.21]

$$z = 0.62616120\ldots \tag{13.33}$$

Constructing finite Markov graphs of increasing length corresponding to $A \to 3.8$ 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:

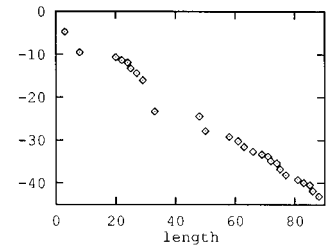



**Figure 13.4:** The logarithm of the difference between the leading zero of the finite polynomial approximations to topological zeta function and our best estimate, as a function of the length for the quadratic map $A = 3.8$.

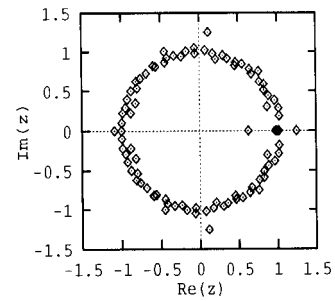



**Figure 13.5:** The 90 zeroes of the characteristic polynomial for the quadratic map $A = 3.8$ approximated by symbolic strings up to length 90. (from ref. [8])

$$h = -\ln 0.62616130424685\ldots = 0.46814726655867\ldots. \tag{13.34}$$

Figure 13.4 illustrates the convergence of the truncation approximations to the topological zeta function as a plot of the logarithm of the difference between the zero of a polynomial and our best estimate (13.34), plotted as a function of the length of the stable periodic orbit. The error of the estimate (13.33) is expected to be of order $z^{29} \approx e^{-14}$ because going from length 28 to a longer truncation yields typically 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 exponent of $-0.47 = -h$, the topological entropy itself. In figure 13.5 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 Markov partitions the topological zeta function has a unit circle as the radius of convergence. The convergence is controlled by the ratio of the leading to the next-to-leading eigenvalues, which is in this case indeed $\lambda_1/\lambda_0 = 1/e^h = e^{-h}$.

## 13.7 Shadowing

The topological zeta function is a pretty function, but the infinite product (13.20) should make you pause. For finite transfer matrices the left hand side is a determinant of a finite matrix, 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 (18.7) of chapter 18; all cycles beyond the fundamental $t_0$

and $t_1$ appear in the shadowing combinations such as

$$t_{s_1 s_2 \cdots s_n} - t_{s_1 s_2 \cdots s_m} t_{s_{m+1} \cdots s_n} \ .$$

For subshifts of finite type such shadowing combinations cancel *exactly*, if we are counting cycles as we do here, or if the dynamics is piecewise linear, as in exercise 17.3. As we have already argued in sect. 1.5.4, for nice hyperbolic flows whose symbolic dynamics is a subshift of finite type, the shadowing combinations *almost* cancel, and the spectral determinant is dominated by the fundamental cycles from (13.13), 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 infinite Markov partition with pruning rules for longer and longer blocks, most of the shadowing combinations still cancel, but the few corresponding to the forbidden blocks do not, leading to a finite radius of convergence for the spectral determinant as in figure 13.5.

One striking aspect of the pruned cycle expansion (13.32) compared to the trace formulas such as (13.7) is that coefficients are not growing exponentially - indeed they all remain of order 1, so instead having a radius of convergence $e^{-h}$, 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 (13.19) increases the *analyticity domain*: the pole in the trace (13.8) at $z = e^{-h}$ is promoted to a smooth zero of the spectral determinant with a larger radius of convergence.

The very sensitive dependence of spectral determinants on whether the symbolic dynamics is or is not a subshift of finite type is the bad news that we should announce already now. If the system is generic and not structurally stable (see sect. 11.3), a smooth parameter variation is in no sense a smooth variation of topological dynamics - infinities of periodic orbits are created or destroyed, Markov graphs go from being finite to infinite and back. That will imply that the global averages that we intend to compute are generically nowhere 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 well ask: What is wrong with computing the entropy from (13.1)? 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 10.13, or the cycle expansion of (13.26) - 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 (13.4). The determinant (13.9) is much smarter, as by construction it encodes the self-similarity of the dynamics, and yields the asymptotic value of $h$ with no need for any finite $n$ extrapolations.

So, the main lesson of learning how to count well, a lesson that will be affirmed over and over, is that while the 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 resumming 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 more convergent and economical to use.

## Résumé

What have we accomplished? We have related the number of topologically distinct paths from "this region" to "that region" in a chaotic system to the leading eigenvalue of the transition matrix $T$. The eigenspectrum of $T$ is given by a certain sum over traces $\mathrm{tr}\, T^n$, and in this way the periodic orbit theory has entered the arena, already at the level of the topological dynamics, the crudest description of dynamics.

The main result of this chapter is the cycle expansion (13.21) of the topological zeta function (i.e., the spectral determinant of the transition matrix):

$$1/\zeta_{\mathrm{top}}(z) \;=\; 1 - \sum_{k=1} \hat{c}_k z^k \,.$$

For subshifts of finite type, the transition matrix is finite, and the topological zeta function is a finite polynomial evaluated by the loop expansion (13.13) 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 smallest zero $z = e^{-h}$. This expression for the entropy is *exact*; in contrast to the definition (13.1), no $n \to \infty$ extrapolations of $\ln K_n/n$ are required.

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 16, a weighted generalization of the topological zeta function.

Contrary to claims one sometimes encounters in the literature, "exponential proliferation of trajectories" is not the problem; what limits the convergence of cycle expansions is the proliferation of the grammar rules, or the "algorithmic complexity," as illustrated by sect. 13.6, and figure 13.5 in particular.

## Commentary

**Remark 13.1** "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 proper invariants one needs to evaluate a supremum over all possible partitions. The key mathematical point that eliminates the need of such search is the existence of *generators*, i.e., partitions that under dynamics are able to probe the whole state space on arbitrarily small scales: more precisely a generator is a finite partition $\Omega = \omega_1 \dots \omega_N$, with the following property: take $\mathcal{M}$ the subalgebra of the state space generated by $\Omega$, and consider the partition built upon all possible intersections of sets $\phi^k(\beta_i)$, where $\phi$ is dynamical evolution, $\beta_i$ is an element of $\mathcal{M}$ and $k$ 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 ref. [1].

**Remark 13.2** Perron-Frobenius matrices. For a proof of Perron theorem on the leading eigenvalue see ref. [22]. Sect. A4.1 of ref. [2] offers a clear discussion of the spectrum of the transition matrix.

**Remark 13.3** Determinant of a graph. Many textbooks offer derivations of the loop expansions of characteristic polynomials for transition matrices and their Markov graphs, see for example refs. [3, 4, 5].

**Remark 13.4** $T$ is not trace class. Note to the erudite reader: the transition matrix $T$ (in the infinite partition limit (13.19)) is *not* trace class.Still the trace is well defined in the $n \to \infty$ limit.

**Remark 13.5** Artin-Mazur zeta functions. Motivated by A. Weil's zeta function for the Frobenius map [8], Artin and Mazur [12] introduced the zeta function (13.21) that counts periodic points for diffeomorphisms (see also ref. [9] for their evaluation for maps of the interval). Smale [10] conjectured rationality of the zeta functions for Axiom A diffeomorphisms, later proved by Guckenheimer [11] and Manning [12]. See remark 17.4 on page 296 for more zeta function history.

**Remark 13.6** Ordering periodic orbit expansions. In sect. 18.5 we will introduce an alternative way of hierarchically organizing cumulant expansions, in which the order is dictated by stability rather than cycle length: such a procedure may be better suited to perform computations when the symbolic dynamics is not well understood.

## Exercises

13.1. **A transition matrix for 3-disk pinball.**

a) Draw the Markov 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 $\mathrm{tr}\, T^n$ to $\mathrm{tr}\, T^{n-1} + \ldots$.)

b) Solve the above difference equation and obtain the number of periodic orbits of length $n$. Compare with table **??**.

c) Find the eigenvalues of the transition matrix $\mathbf{T}$ 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 $\{0, 1\}$.

13.2. **Sum of $A_{ij}$ is like a trace.** Let $A$ be a matrix with eigenvalues $\lambda_k$. Show that

$$\Gamma_n = \sum_{i,j} [A^n]_{ij} = \sum_k c_k \lambda_k^n .$$

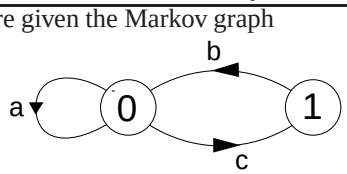
(a) Use this to show that $\ln|\mathrm{tr}\, A^n|$ and $\ln|\Gamma_n|$ have the same asymptotic behavior as $n \to \infty$, i.e., their ratio converges to one.

(b) Do eigenvalues $\lambda_k$ need to be distinct, $\lambda_k \neq \lambda_l$ for $k \neq l$?

13.3. **Loop expansions.** Prove by induction the sign rule in the determinant expansion (13.13):

$$\det(1 - z\mathbf{T}) = \sum_{k \geq 0} \sum_{p_1 + \cdots + p_k} (-1)^k t_{p_1} t_{p_2} \cdots t_{p_k} .$$

13.4. **Transition matrix and cycle counting.** Suppose you are given the Markov 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.

a) Walks on the graph are given the 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:

$$T = \begin{bmatrix} a & b \\ c & 0 \end{bmatrix} .$$

b) Enumerate all the walks of length three on the Markov 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 $T^n_{ij}$ is the number of walks from point $i$ to point $j$ in $n$ steps. (Hint: one might use the method of induction.)

d) Try to estimate the number $N(n)$ of walks of length $n$ for this simple Markov graph.

e) The topological entropy $h$ measures the rate of exponential growth of the total number of walks $N(n)$ as a function of $n$. What is the topological entropy for this Markov graph?

13.5. **3-disk prime cycle counting.** A *prime* cycle $p$ of length $n_p$ 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, $\cdots$ prime cycles of length 2, 3, 4, 5, 6, $\cdots$.

13.6. **"Golden mean" pruned map.** Continuation of exercise 10.6: Show that the total number of periodic orbits of length $n$ for the "golden mean" tent map is

$$\frac{(1+\sqrt{5})^n + (1-\sqrt{5})^n}{2^n} .$$

For continuation, see exercise 17.2. See also exercise 13.8.

13.7. **Alphabet {0,1}, prune $\overline{00}$.** The Markov diagram figure 10.13 (b) implements this pruning rule. The pruning rule 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 with binary alphabet {1,2}. The cycle expansion (13.13) becomes

$$\begin{aligned} 1/\zeta &= (1-t_1)(1-t_2)(1-t_{12})(1-t_{112}) \\ &= 1 - t_1 - t_2 - (t_{12} - t_1 t_2) \\ &\quad -(t_{112} - t_{12}t_1) - (t_{122} - t_{12}t_2) \ldots \end{aligned} \tag{13.35}$$

In the original binary alphabet this corresponds to:

$$\begin{aligned} 1/\zeta &= 1 - t_1 - t_{10} - (t_{110} - t_1 t_{10}) \\ &\quad -(t_{1110} - t_{110}t_1) - (t_{11010} - t_{110}t_{10}) \end{aligned} \tag{13.36}$$

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 13.6.

13.8. **A unimodal map example.** Consider a unimodal map, this Figure (a):

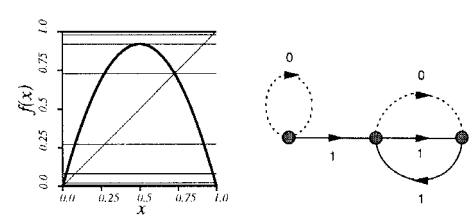


**Figure:** *(a) A unimodal map for which the critical point maps into the right hand fixed point in three iterations, and (b) the corresponding Markov graph (K.T. Hansen).* for which the critical point maps into the right hand fixed point in three iterations, $S^+ = 100\overline{1}$. Show that the admissible itineraries are generated by the Markov graph of the Figure (b).

(Kai T. Hansen)

13.9. **Glitches in shadowing.**** Note that the combination $t_{00011}$ minus the "shadow" $t_0 t_{0011}$ in (13.17) cancels exactly, and does not contribute to the topological zeta function (13.18). Are you able to construct a smaller Markov graph than figure 13.3 (e)?

13.10. **Whence Möbius function?** To understand where the Möbius function comes from consider the function

$$f(n) = \sum_{d|n} g(d) \tag{13.37}$$

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*

$$g(n) = \sum_{d|n} \mu(n/d) f(d) \tag{13.38}$$

13.11. **Counting prime binary cycles.** In order to get comfortable with Möbius inversion reproduce the results of the second column of table **??**.

Write a program that determines the number of prime cycles of length $n$. You might want to have this program later on to be sure that you have missed no 3-pinball prime cycles.

13.12. **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 (10.6) 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 (13.5) collapse to $tr(T^k) = (t_0 + t_1)^k$, and $1/\zeta$ is simply

$$\prod_p \left(1 - t_p\right) = 1 - t_0 - t_1 \tag{13.39}$$

Substituting (13.39) into the identity

$$\prod_p \left(1 + t_p\right) = \prod_p \frac{1 - {t_p}^2}{1 - t_p}$$

we obtain

$$\begin{aligned}\prod_p \left(1 + t_p\right) &= \frac{1 - t_0^2 - t_1^2}{1 - t_0 - t_1} \\ &= 1 + t_0 + t_1 + \frac{2t_0 t_1}{1 - t_0 - t_1} \\ &= 1 + t_0 + t_1 \\ &\quad + \sum_{n=2}^{\infty} \sum_{k=1}^{n-1} 2 \binom{n-2}{k-1} t_0^k t_1^{n-k}.\end{aligned}$$

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 $M_{n,k}$ $(n = 1, 2, \ldots;\ k = \{0, 1\}$ for $n = 1$; $k = 1, \ldots, n-1$ for $n \geq 2)$ the number of prime $n$-cycles whose labels contain $k$ zeros. Show that

$$\begin{aligned} M_{1,0} &= M_{1,1} = 1\,, \quad n \geq 2\,, k = 1, \ldots, n-1 \\ nM_{n,k} &= \sum_{m \left| \frac{n}{k} \right.} \mu(m) \binom{n/m}{k/m} \end{aligned}$$

where the sum is over all $m$ which divide both $n$ and $k$. (Continued as exercise 18.7.)

13.13. **Logarithmic periodicity of** $\ln N_n{}^*$**.** Plot $\ln N_n - nh$ for a system with a nontrivial finite Markov graph. Do you see any periodicity? If yes, why?

13.14. **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

$$\begin{aligned} 1/\zeta_{\text{top}}^{4-disk} &= (1 - 3z) \frac{(1 - z^2)^6}{(1 - z)^3 (1 - z^2)^3} \\ &= (1 - 3z)(1 + z)^3 \\ &= 1 - 6z^2 - 8z^3 - 3z^4\ . \end{aligned} \tag{13.40}$$

13.15. $N$**-disk pinball topological zeta function.** Show that for an $N$-disk pinball, the topological zeta function is given by

$$\begin{aligned} 1/\zeta_{\text{top}}^{N-disk} &= (1 - (N-1)z) \times \\ &\quad \frac{(1 - z^2)^{N(N-1)/2}}{(1 - z)^{N-1} (1 - z^2)^{(N-1)(N-2)/2}} \\ &= (1 - (N-1)z)\,(1 + z)^{N-1}\ . \end{aligned} \tag{13.41}$$

The topological zeta function has a root $z^{-1} = N - 1$, as we already know it should from (13.29) or (13.15). We shall see in sect. 19.4 that the other roots reflect the symmetry factorizations of zeta functions.

13.16. **Alphabet** $\{a, b, c\}$**, prune** $_ab_$**.** The pruning rule implies that any string of "b"s must be preceeded by a "c"; so one possible alphabet is $\{a, cb^k;\ \overline{b}\}$, k=0,1,2. . ..

As the rule does not prune the fixed point $\overline{b}$, it is explicitly included in the list. The cycle expansion (13.13) becomes

$$\begin{aligned} 1/\zeta &= (1-t_a)(1-t_b)(1-t_c)\times \\ &\quad (1-t_{cb})(1-t_{ac})(1-t_{cbb})\ldots \\ &= 1-t_a-t_b-t_c+t_at_b-(t_{cb}-t_ct_b) \\ &\quad -(t_{ac}-t_at_c)-(t_{cbb}-t_{cb}t_b)\ldots \end{aligned}$$

The effect of the $_ab_$ pruning is essentially to unbalance the 2 cycle curvature $t_{ab}-t_at_b$; the remainder of the cycle expansion retains the curvature form.

13.17. **Alphabet {0,1}, prune $n$ repeats** of "0" $_000\ldots00_$. This is equivalent to the $n$ symbol alphabet {1, 2, ..., $n$} unrestricted symbolic dynamics, with symbols corresponding to the possible 10...00 block lengths: 2=10, 3=100, ..., $n$=100...00. The cycle expansion (13.13) becomes

$$1/\zeta = 1-t_1-t_2\ldots-t_n-(t_{12}-t_1t_2)\ldots-(t_{1n}-t_1t_n)\ldots \quad (13.42)$$

13.18. **Alphabet {0,1}, prune $_1000_$, $_00100_$, $_01100_$** Show that the topological zeta function is given by

$$1/\zeta = (1-t_0)(1-t_1-t_2-t_{23}-t_{113}) \quad (13.43)$$

with the unrestricted 4-letter alphabet {1, 2, $\underline{23}$, $\underline{113}$}. Here 2, 3, refer to 10, 100 respectively, as in exercise 13.17.

13.19. **Alphabet {0,1}, prune $_1000_$, $_00100_$, $_01100_$, $_10011_$.** 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 13.18) to the alphabet {$\underline{21^k}$, $\underline{23}$, $\underline{21^k113}$; $\overline{1}$, $\overline{0}$}, and the cycle expansion

$$1/\zeta = (1-t_0)(1-t_1-t_2-t_{23}+t_1t_{23}-t_{2113}) \quad (13.44)$$

Note that this says that 1, 23, 2, 2113 are the fundamental cycles; not all cycles up to length 7 are needed, only 2113.

(b) Show that the topological zeta function is

$$1/\zeta_{\rm top} = (1-z)(1-z-z^2-z^5+z^6-z^7) \quad (13.45)$$

and check that it yields the exact value of the entropy $h = 0.522737642\ldots$.

13.20. **Topological zeta function for alphabet {0,1}, prune $_1000_$, $_00100_$, $_01100_$** (continuation of exercise 11.9) Show that topological zeta function is

$$1/\zeta = (1-t_0)(1-t_1-t_2-t_{23}-t_{113}) \quad (13.46)$$

for unrestricted 4-letter alphabet {1, 2, $\underline{23}$, $\underline{113}$}.

13.21. **Alphabet {0,1}, prune only the fixed point $\overline{0}$.** This is equivalent to the *infinite* alphabet {1, 2, 3, 4, ...} unrestricted symbolic dynamics. The prime cycles are labeled by all non-repeating sequences of integers, ordered lexically: $t_n, n>0$; $t_{mn}, t_{mmn}, \ldots, n>m>0$; $t_{mnr}, r>n>m>0, \ldots$ (see sect. 23.3). Now the number of fundamental cycles is infinite as well:

$$\begin{aligned} 1/\zeta &= 1-\sum_{n>0} t_n - \sum_{n>m>0}(t_{mn}-t_nt_m) \\ &\quad -\sum_{n>m>0}(t_{mmn}-t_mt_{mn}) \\ &\quad -\sum_{n>m>0}(t_{mnn}-t_{mn}t_n) \qquad (13.47) \\ &\quad -\sum_{r>n>m>0}(t_{mnr}+t_{mrn}-t_{mn}t_r \\ &\quad -t_{mr}t_n-t_mt_{nr}+t_mt_nt_r)\cdots \qquad (13.48) \end{aligned}$$

As shown in table **??**, this grammar plays an important role in description of fixed points of marginal stability.

## References


[13.1] V.I. Arnold and A. Avez, "Ergodic Problems of Classical Mechanics," *Addison-Wesley, Redwood City* (1989).

[13.2] J. Zinn-Justin, "Quantum Field Theory and Critical Phenomena," *Clarendon Press, Oxford* (1996).

[13.3] A. Salomaa, "Formal Languages," *Academic Press, San Diego* (1973).

[13.4] J.E. Hopcroft and J.D. Ullman, "Introduction to Automata Theory, Languages and Computation," *Addison-Wesley, Reading Ma* (1979).

[13.5] D.M. Cvektović, M. Doob and H. Sachs, "Spectra of Graphs," *Academic Press, New York* (1980).

[13.6] J. Riordan, *An Introduction to Combinatorial Analysis*, ( Wiley, New York 1958) ; E.N. Gilbert and J. Riordan, *Illinois J.Math* **5**, 657 (1961).

[13.7] K.M. Brucks, *Adv. Appl. Math.* **8**, 434 (1987).

[13.8] A. Weil, *Bull.Am.Math.Soc.* **55**, 497 (1949).

[13.9] J. Milnor and W. Thurston, "On iterated maps of the interval," in A. Dold and B. Eckmann, eds., *Dynamical Systems, Proceedings, U. of Maryland 1986-87, Lec. Notes in Math.* **1342**, 465 (Springer, Berlin 1988).

[13.10] S. Smale, *Ann. Math.*, **74**, 199 (1961).

[13.11] J. Guckenheimer, *Invent. Math.* **39**, 165 (1977).

[13.12] A. Manning, *Bull. London Math. Soc.* **3**, 215 (1971).

[13.13] A.L. Kholodenko, "Designing new apartment buildings for strings and conformal field theories. First steps," `arXiv:hep-th/0312294`

# Chapter 14

# 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, 7 and 8 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 15 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 16 to 18 show how to compute these without ever having to compute a natural" invariant densities $\rho_0$.

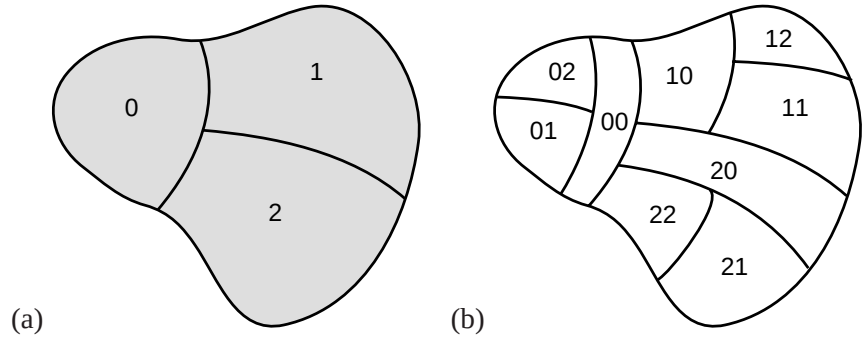



**Figure 14.1:** (a) First level of partitioning: A coarse partition of $\mathcal{M}$ into regions $\mathcal{M}_0$, $\mathcal{M}_1$, and $\mathcal{M}_2$. (b) $n = 2$ level of partitioning: A refinement of the above partition, with each region $\mathcal{M}_i$ subdivided into $\mathcal{M}_{i0}$, $\mathcal{M}_{i1}$, and $\mathcal{M}_{i2}$.

## 14.1 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, ..., $n$, ... of increasingly refined partitions of state space, figure 14.1, into regions $\mathcal{M}_i$ defined by the characteristic function

$$\chi_i(x) = \begin{cases} 1 & \text{if } x \in \mathcal{M}_i\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{14.1}$$

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:

$$\Delta\mu_i = \int_{\mathcal{M}} d\mu(x)\chi_i(x) = \int_{\mathcal{M}_i} d\mu(x) = \int_{\mathcal{M}_i} dx\,\rho(x)\,. \tag{14.2}$$

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

$$\sum_i^{(n)} \Delta\mu_i = 1\,, \tag{14.3}$$

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 $\Delta\mu_i = |\mathcal{M}_i|\rho(x_i)\,,\ x_i \in \mathcal{M}_i$, with normalization

$$\int_{\mathcal{M}} dx\,\rho(x) = 1\,. \tag{14.4}$$

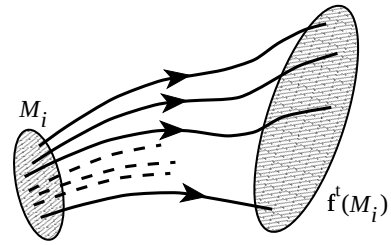



**Figure 14.2:** The evolution rule $f^t$ can be used to map a region $\mathcal{M}_i$ of the state space into the region $f^t(\mathcal{M}_i)$.

Here $|\mathcal{M}_i|$ is the volume of region $\mathcal{M}_i$, and all $|\mathcal{M}_i| \to 0$ as $n \to \infty$.

So far, any arbitrary sequence of partitions will do. What are intelligent ways of partitioning state space? We already know the answer from chapter 10, but let us anyway develope some intuition about how the dynamics transports densities.

[chapter 10]

## 14.2 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 $\mathcal{M}_i$ at time $t = 0$. As the flow evolves, this region is carried into $f^t(\mathcal{M}_i)$, as in figure 14.2. No trajectory is created or destroyed, so the conservation of representative points requires that

$$\int_{f^t(\mathcal{M}_i)} dx\, \rho(x,t) = \int_{\mathcal{M}_i} dx_0\, \rho(x_0,0)\,.$$

Transform the integration variable in the expression on the left hand side to the initial points $x_0 = f^{-t}(x)$,

$$\int_{\mathcal{M}_i} dx_0\, \rho(f^t(x_0),t)\left|\det J^t(x_0)\right| = \int_{\mathcal{M}_i} dx_0\, \rho(x_0,0)\,.$$

The density changes with time as the inverse of the Jacobian (4.46)

$$\rho(x,t) = \frac{\rho(x_0,0)}{|\det J^t(x_0)|}\,, \qquad x = f^t(x_0)\,, \tag{14.5}$$

which makes sense: the density varies inversely with the infinitesimal volume occupied by the trajectories of the flow.

The relation (14.5) is linear in $\rho$, so the manner in which a flow transports densities may be recast into the language of operators, by writing

[exercise 14.1]

$$\rho(x,t) = \left(\mathcal{L}^t \circ \rho\right)(x) = \int_{\mathcal{M}} dx_0\, \delta\Big(x - f^t(x_0)\Big)\rho(x_0,0)\,. \tag{14.6}$$

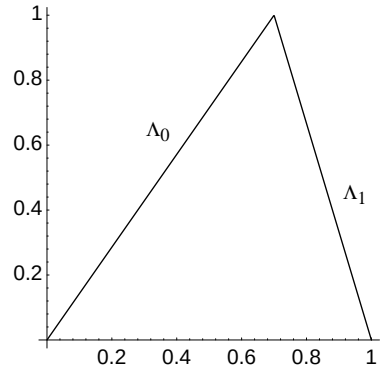


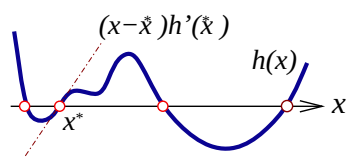

**Figure 14.3:** A piecewise-linear skew "Ulam tent" map (14.11) ($\Lambda_0 = 4/3$, $\Lambda_1 = -4$).

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:

$$\int dx\,\delta(h(x)) = \sum_{\{x:h(x)=0\}} \frac{1}{|h'(x)|}\,, \tag{14.7}$$

and in $d$ dimensions the denominator is replaced by

$$\int dx\,\delta(h(x)) \;=\; \tag{14.8}$$




$$= \sum_j \int_{\mathcal{M}_j} dx\,\delta(h(x)) \;=\; \sum_{\{x:h(x)=0\}} \frac{1}{\left|\det \frac{\partial h(x)}{\partial x}\right|}\,.$$

Now you can check that (14.6) is just a rewrite of (14.5): [exercise 14.2]

$$\begin{aligned}\left(\mathcal{L}^t \circ \rho\right)(x) &= \sum_{x_0=f^{-t}(x)} \frac{\rho(x_0)}{|f^{t\prime}(x_0)|} && \text{(1-dimensional)} \\ &= \sum_{x_0=f^{-t}(x)} \frac{\rho(x_0)}{|\det J^t(x_0)|} && (d\text{-dimensional})\,. \end{aligned} \tag{14.9}$$

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, to be discussed briefly in the next example, and in more detail in sect. 10.2.1.

We shall refer to the kernel of (14.6) as the *Perron-Frobenius operator*: [exercise 14.3] [example 21.7]

$$\mathcal{L}^t(x,y) = \delta\Big(x - f^t(y)\Big)\,. \tag{14.10}$$

If you do not like the word "kernel" you might prefer to think of $\mathcal{L}^t(x,y)$ as a matrix with indices $x, y$, and index summation in matrix multiplication replaced by an integral over $y$, $(\mathcal{L}^t \circ \rho)(x) = \int dy\, \mathcal{L}^t(x,y)\rho(y)$. The Perron-Frobenius operator assembles the density $\rho(x,t)$ at time $t$ by going back in time to the density $\rho(x_0, 0)$ at time $t = 0$. [remark 17.4]

**Example 14.1** ***Perron-Frobenius operator for a piecewise-linear map:*** *Assume the expanding 1-d map $f(x)$ of figure 14.3, a piecewise-linear 2–branch map with slopes $\Lambda_0 > 1$ and $\Lambda_1 = -\Lambda_0/(\Lambda_0 - 1) < -1$ :* [exercise 14.7]

$$f(x) = \begin{cases} f_0(x) = \Lambda_0 x\,, & x \in \mathcal{M}_0 = [0, 1/\Lambda_0) \\ f_1(x) = \Lambda_1(1-x)\,, & x \in \mathcal{M}_1 = (1/\Lambda_0, 1]\,. \end{cases} \tag{14.11}$$

*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*

$$\rho(x) = \begin{cases} \rho_0 & \text{if } x \in \mathcal{M}_0 \\ \rho_1 & \text{if } x \in \mathcal{M}_1 \end{cases}\,. \tag{14.12}$$

*As can be easily checked using (14.9), the Perron-Frobenius operator acts on this piecewise constant function as a [2×2] Markov matrix $\mathbf{L}$ with matrix elements* [exercise 14.1] [exercise 14.5]

$$\begin{pmatrix} \rho_0 \\ \rho_1 \end{pmatrix} \to \mathbf{L}\rho = \begin{pmatrix} \frac{1}{|\Lambda_0|} & \frac{1}{|\Lambda_1|} \\ \frac{1}{|\Lambda_0|} & \frac{1}{|\Lambda_1|} \end{pmatrix} \begin{pmatrix} \rho_0 \\ \rho_1 \end{pmatrix}, \tag{14.13}$$

*stretching both $\rho_0$ and $\rho_1$ over the whole unit interval $\Lambda$. 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 $\rho$ to the space of piecewise constant functions. The example gives a flavor of the enterprize upon which we are about to embark in this book, 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 15.2.)*

## 14.3 Why not just leave it to a computer?

(R. Artuso and P. Cvitanović)

To a student with a practical bent the above Example 14.1 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 $\rho$ is crucial, and the physically motivated choice is a space of smooth functions, rather than the space of piecewise constant functions. [chapter 21]

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 21). 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 14.4, is to partition the state space $\mathcal{M}$ with a non-overlapping collection of sets $\mathcal{M}_i$, $i = 1, \ldots, N$, and to consider piecewise constant densities (14.2), constant on each $\mathcal{M}_i$:

$$\rho(x) = \sum_{i=1}^{N} \rho_i \frac{\chi_i(x)}{|\mathcal{M}_i|}$$

where $\chi_i(x)$ is the characteristic function (14.1) of the set $\mathcal{M}_i$. The density $\rho_i$ at a given instant is related to the densities at the previous step in time by the action of the Perron-Frobenius operator, as in (14.6):

$$\begin{aligned} \rho'_j &= \int_{\mathcal{M}} dy\, \chi_j(y) \rho'(y) = \int_{\mathcal{M}} dx\, dy\, \chi_j(y)\, \delta(y - f(x))\, \rho(x) \\ &= \sum_{i=1}^{N} \rho_i \frac{|\mathcal{M}_i \cap f^{-1}(\mathcal{M}_j)|}{|\mathcal{M}_i|} . \end{aligned}$$

In this way

$$\mathbf{L}_{ij} = \frac{|\mathcal{M}_i \cap f^{-1}(\mathcal{M}_j)|}{|\mathcal{M}_i|} , \quad \rho' = \rho \mathbf{L} \tag{14.14}$$

is a matrix approximation to the Perron-Frobenius operator, and its leading left eigenvector is a piecewise constant approximation to the invariant measure. It is an old idea of Ulam that such an approximation for the Perron-Frobenius operator is a meaningful one.

[remark 14.3]

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 10, 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 1.11.

Furthermore, a piecewise constant $\rho$ belongs to an unphysical function space, and with such approximations one is plagued by numerical artifacts such as spurious eigenvalues. In chapter 21 we shall employ a more refined approach to

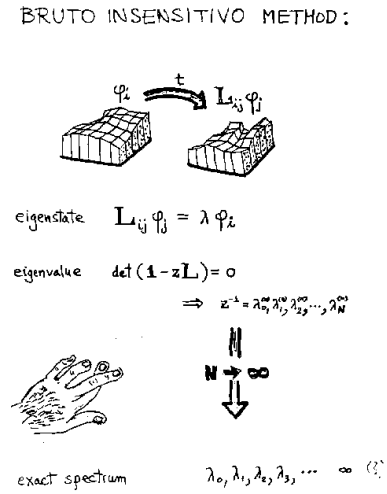



**Figure 14.4:** State space discretization approach to computing averages.

extracting spectra, by expanding the initial and final densities $\rho$, $\rho'$ in some basis $\varphi_0$, $\varphi_1$, $\varphi_2, \cdots$ (orthogonal polynomials, let us say), and replacing $\mathcal{L}(y,x)$ by its $\varphi_\alpha$ 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 21]

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 14.5, 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 $\varphi_\alpha$ 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. 1

## 14.4 Invariant measures

A *stationary* or *invariant density* is a density left unchanged by the flow

$$\rho(x,t) = \rho(x,0) = \rho(x)\,. \tag{14.15}$$

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 $\rho$ densities) of the Perron-Frobenius operator (14.10), with the unit eigenvalue:

[exercise 14.3]

$$\mathcal{L}^t\rho(x) = \int_{\mathcal{M}} dy\, \delta(x - f^t(y))\rho(y) = \rho(x). \tag{14.16}$$

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 (14.16). 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 $f$,

$$\rho_0(x) = \lim_{t\to\infty} \int_{\mathcal{M}} dy\, \delta(x - f^t(y))\rho(y, 0)\,, \qquad \int_{\mathcal{M}} dy\, \rho(y, 0) = 1\,. \tag{14.17}$$

Intuitively, the "natural" measure should be the measure that is the least sensitive to the (in practice unavoidable) external noise, no matter how weak.

### 14.4.1 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 14.5, the long time evolution of many "typical" initial conditions leads to the same asymptotic distribution. Hence the *natural* (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 14.8]
[exercise 14.9]

$$\overline{\rho}_{x_0}(y) \quad = \quad \begin{cases} \lim_{t\to\infty} \frac{1}{t}\int_0^t d\tau\, \delta(y - f^\tau(x_0)) & \text{flows} \\ \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \delta\Big(y - f^k(x_0)\Big) & \text{maps}\,, \end{cases} \tag{14.18}$$

where $x_0$ is a generic initial point. Generated by the action of $f$, the natural measure satisfies the stationarity condition (14.16) 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 (14.18) over the $\mathcal{M}_i$ region

$$\Delta\overline{\mu}_i = \lim_{t\to\infty} \frac{t_i}{t}\,, \tag{14.19}$$

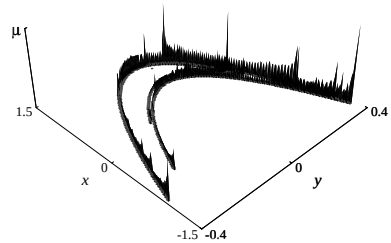



**Figure 14.5:** Natural measure (14.19) for the Hénon map (3.18) strange attractor at parameter values $(a, b) = (1.4, 0.3)$. See figure 3.9 for a sketch of the attractor without the natural measure binning. (Courtesy of J.-P. Eckmann)

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. 15.1.

The *space average* of the observable $a$ with respect to a measure $\rho$ is given by the $d$-dimensional integral over the state space $\mathcal{M}$:

$$\begin{aligned} \langle a \rangle_\rho &= \frac{1}{|\rho_{\mathcal{M}}|} \int_{\mathcal{M}} dx\, \rho(x) a(x) \\ |\rho_{\mathcal{M}}| &= \int_{\mathcal{M}} dx\, \rho(x) = \text{mass in } \mathcal{M}\,. \end{aligned} \tag{14.20}$$

For now we assume that the state space $\mathcal{M}$ has a finite dimension and a finite volume. By definition, $\langle a \rangle_\rho$ is a function(al) of $\rho$. 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 (14.18) into (14.20), we see that the natural measure corresponds to a *time average* of the observable $a$ along a trajectory of the initial point $x_0$,

$$\overline{a_{x_0}} = \lim_{t\to\infty} \frac{1}{t} \int_0^t d\tau\, a(f^\tau(x_0))\,. \tag{14.21}$$

Analysis of the above asymptotic time limit is the central problem of ergodic theory. The *Birkhoff ergodic theorem* asserts that if a natural measure $\rho$ exists, the limit $\overline{a(x_0)}$ for the time average (14.21) exists for all initial $x_0$. As we shall not rely on this result in what follows we forgo a proof here. Furthermore, if the dynamical system is *ergodic*, the time average tends to the space average [remark 14.1] [appendix A]

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t d\tau\, a(f^\tau(x_0)) = \langle a \rangle \tag{14.22}$$

for "almost all" initial $x_0$. By "almost all" we mean that the time average is independent of the initial point apart from a set of $\rho$-measure zero.

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,

$$\lim_{t\to\infty} \left\langle a(x)b(f^t(x)) \right\rangle = \langle a \rangle \langle b \rangle \ , \tag{14.23}$$

[section 20.4]

the dynamical system is said to be *mixing*.

**Example 14.2** ***The Hénon attractor natural measure:*** *A numerical calculation of the natural measure (14.19) for the Hénon attractor (3.18) is given by the histogram in figure 14.5. The state space is partitioned into many equal-size areas $\mathcal{M}_i$, and the coarse grained measure (14.19) 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.*

If an invariant measure is quite singular (for instance a Dirac $\delta$ concentrated on a fixed point or a cycle), its existence is most likely of no physical import; no smooth initial density will converge to this measure if its neighborhood is repelling. In practice the average (14.18) 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 14.5, is "strange," or merely a transient to a very long stable cycle.

[exercise 15.1]

### 14.4.2 Determinism vs. stochasticity

While dynamics can lead to very singular $\rho$'s, in any physical setting we cannot do better than to measure $\rho$ averaged over some region $\mathcal{M}_i$; 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 became 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 17]

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 this book. 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 $\Delta\rho_i$.

## 14.5 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 $\delta\tau$:

$$\begin{aligned}
\mathcal{L}^{\delta\tau}\rho(y) &= \int_{\mathcal{M}} dx\, \delta\Big(y - f^{\delta\tau}(x)\Big)\rho(x) \\
&= \int_{\mathcal{M}} dx\, \delta(y - x - \delta\tau v(x))\,\rho(x) \\
&= \frac{\rho(y - \delta\tau v(y))}{\left|\det\left(1 + \delta\tau\frac{\partial v(y)}{\partial x}\right)\right|} = \frac{\rho(y) - \delta\tau v_i(y)\partial_i\rho(y)}{1 + \delta\tau\sum_{i=1}^d \partial_i v_i(y)} \\
\rho(x,\delta\tau) &= \rho(x,0) - \delta\tau\,. \qquad (14.24)
\end{aligned}$$

Here we have used the infinitesimal form of the flow (2.6), the Dirac delta Jacobian (14.9), and the ln det = 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 $\rho$ under the infinitesimal action of the Perron-Frobenius operator is nothing but the *continuity equation* for the density: [exercise 4.1]

$$\partial_t\rho + \partial\cdot(\rho v) = 0\,. \qquad (14.25)$$

The family of Perron-Frobenius operators operators $\{\mathcal{L}^t\}_{t\in\mathbb{R}_+}$ forms a semigroup parameterize by time

(a) $\mathcal{L}^0 = I$

(b) $\mathcal{L}^t\mathcal{L}^{t'} = \mathcal{L}^{t+t'} \qquad t,t' \geq 0 \qquad$ (semigroup property) .

From (14.24), time evolution by an infinitesimal step $\delta\tau$ forward in time is generated by

$$\mathcal{A}\rho(x) = +\lim_{\delta\tau\to 0^+}\frac{1}{\delta\tau}\Big(\mathcal{L}^{\delta\tau} - I\Big)\rho(x) = -\partial_i(v_i(x)\rho(x))\,. \qquad (14.26)$$

We shall refer to

$$\mathcal{A} = -\partial\cdot v + \sum_i^d v_i(x)\partial_i \qquad (14.27)$$

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 (14.26) is the definition of time derivative, so the evolution equation for $\rho(x)$ is

$$\left(\frac{\partial}{\partial t} - \mathcal{A}\right)\rho(x) = 0\,. \qquad (14.28)$$

The finite time Perron-Frobenius operator (14.10) can be formally expressed by exponentiating the time evolution generator $\mathcal{A}$ as

$$\mathcal{L}^t = e^{t\mathcal{A}} . \tag{14.29}$$

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):

[exercise 14.10]

$$e^{-tv\frac{\partial}{\partial x}} a(x) = a(x - tv) . \tag{14.30}$$

### 14.5.1 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 $M > 0$, $\beta \geq 0$ such that $\|\mathcal{L}^t\| \leq Me^{t\beta}$ 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 $e^{t\beta}$, a reasonable expectation in the light of the simplest example studied so far, the exact escape rate (15.20). If that is so, multiplying $\mathcal{L}^t$ by $e^{-t\beta}$ we construct a new operator $e^{-t\beta}\mathcal{L}^t = e^{t(\mathcal{A}-\beta)}$ which decays exponentially for large $t$, $\|e^{t(\mathcal{A}-\beta)}\| \leq M$. We say that $e^{-t\beta}\mathcal{L}^t$ is an element of a *bounded* semigroup with generator $\mathcal{A} - \beta I$. Given this bound, it follows by the Laplace transform

$$\int_0^\infty dt\, e^{-st}\mathcal{L}^t = \frac{1}{s - \mathcal{A}} , \qquad \mathrm{Re}\, s > \beta , \tag{14.31}$$

that the *resolvent* operator $(s - \mathcal{A})^{-1}$ is bounded ("resolvent" = able to cause separation into constituents)

$$\left\| \frac{1}{s - \mathcal{A}} \right\| \leq \int_0^\infty dt\, e^{-st} M e^{t\beta} = \frac{M}{s - \beta} .$$

If one is interested in the spectrum of $\mathcal{L}$, as we will be, the resolvent operator is a natural object to study; it has no time dependence, and it is bounded. 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 (14.29) into the resolvent form (14.31) and enables us to study its spectrum.

## 14.6 Liouville operator

A case of special interest is the Hamiltonian or symplectic flow defined by Hamilton's equations of motion (7.1). A reader versed in quantum mechanics will have observed by now that with replacement $\mathcal{A} \to -\frac{i}{\hbar}\hat{H}$, where $\hat{H}$ is the quantum Hamiltonian operator, (14.28) looks rather like the time dependent Schrödinger equation, so this is probably the right moment to figure out what all this means in the case of Hamiltonian flows.

The Hamilton's evolution equations (7.1) for any time-independent quantity $Q = Q(q, p)$ are given by

$$\frac{dQ}{dt} = \frac{\partial Q}{\partial q_i}\frac{dq_i}{dt} + \frac{\partial Q}{\partial p_i}\frac{dp_i}{dt} = \frac{\partial H}{\partial p_i}\frac{\partial Q}{\partial q_i} - \frac{\partial Q}{\partial p_i}\frac{\partial H}{\partial q_i}\,. \tag{14.32}$$

As equations with this structure arise frequently for symplectic flows, it is convenient to introduce a notation for them, the *Poisson bracket* [remark 14.4]

$$\{A, B\} = \frac{\partial A}{\partial p_i}\frac{\partial B}{\partial q_i} - \frac{\partial A}{\partial q_i}\frac{\partial B}{\partial p_i}\,. \tag{14.33}$$

In terms of Poisson brackets the time evolution equation (14.32) takes the compact form

$$\frac{dQ}{dt} = \{H, Q\}\,. \tag{14.34}$$

The full state space flow velocity is $\dot{x} = v = (\dot{q}, \dot{p})$, where the dot signifies time derivative.

The discussion of sect. 14.5 applies to any deterministic flow. If the density itself is a material invariant, combining

$$\partial_t I + v \cdot \partial I = 0\,.$$

and (14.25) 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 of sect. 7.2. For incompressible flows the continuity equation (14.25) becomes a statement of conservation of the state space volume (see sect. 7.2), or the *Liouville theorem*

$$\partial_t \rho + v_i \partial_i \rho = 0\,. \tag{14.35}$$

Hamilton's equations (7.1) imply that the flow is incompressible, $\partial_i v_i = 0$, so for Hamiltonian flows the equation for $\rho$ reduces to the *continuity equation* for the phase space density:

$$\partial_t \rho + \partial_i(\rho v_i) = 0\,, \qquad i = 1, 2 \ldots, D\,. \tag{14.36}$$

Consider the evolution of the phase space density $\rho$ 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 (7.1) we obtain the *Liouville equation*, a special case of (14.28):

$$\frac{\partial}{\partial t}\rho(q,p,t) = -\mathcal{A}\rho(q,p,t) = \{H,\rho(q,p,t)\}\,, \tag{14.37}$$

where $\{\ ,\ \}$ is the Poisson bracket (14.33). The generator of the flow (14.27) is in this case a generator of infinitesimal symplectic transformations,

$$\mathcal{A} = \dot{q}_i\frac{\partial}{\partial q_i} + \dot{p}_i\frac{\partial}{\partial p_i} = \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q_i} - \frac{\partial H}{\partial q_i}\frac{\partial}{\partial p_i}\,. \tag{14.38}$$

For example, for separable Hamiltonians of form $H = p^2/2m+V(q)$, the equations of motion are

$$\dot{q}_i = \frac{p_i}{m}\,, \qquad \dot{p}_i = -\frac{\partial V(q)}{\partial q_i}\,. \tag{14.39}$$

and the action of the generator

[exercise 14.11]

$$\mathcal{A} = -\frac{p_i}{m}\frac{\partial}{\partial q_i} + \partial_i V(q)\frac{\partial}{\partial p_i}\,. \tag{14.40}$$

can be interpreted as a translation (14.30) in configuration space, followed by acceleration by force $\partial V(q)$ in the momentum space.

The time evolution generator (14.27) 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 $10^{23}$ 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.

## 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 accurately 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.

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*,

$$\rho(x,t) = \left(\mathcal{L}^t \circ \rho\right)(x)\,,$$

and in a differential formulation they satisfy the 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, classical 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. For long times the dynamics is described in terms of stationary measures, i.e., fixed points of the appropriate evolution operators. Both in classical and quantum mechanics one has a choice of implementing dynamical evolution on densities ("Schrödinger picture," sect.14.5) or on observables ("Heisenberg picture," sect. 15.2 and chapter 16).

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 classical evolution operators are not unitary, their eigenstates 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 $\delta$-functions in state space. The trick is that each comes with an interval, its neighborhood – cycle points only serve to pin these intervals, just as the millimeter marks on a measuring rod partition continuum into intervals.

## Commentary

**Remark 14.1** Ergodic theory: An overview of ergodic theory is outside the scope of this book: the interested reader may find it useful to consult ref. [1]. The existence of

time average (14.21) is the basic result of ergodic theory, known as the Birkhoff theorem, see for example refs. [1, 22], or the statement of theorem 7.3.1 in ref. [8]. The natural measure (14.19) of sect. 14.4.1 is often referred to as the SRB or Sinai-Ruelle-Bowen measure [26, 24, 28].

**Remark 14.2** Time evolution as a Lie group: Time evolution of sect. 14.5 is an example of a 1-parameter Lie group. Consult, for example, chapter 2. of ref. [9] for a clear and pedagogical introduction to Lie groups of transformations. For a discussion of the bounded semigroups of page 246 see, for example, Marsden and Hughes [2].

**Remark 14.3** Discretization of the Perron-Frobenius operator operator It is an old idea of Ulam [12] that such an approximation for the Perron-Frobenius operator is a meaningful one. The piecewise-linear approximation of the Perron-Frobenius operator (14.14) has been shown to reproduce the spectrum for expanding maps, once finer and finer Markov partitions are used [13, 17, 14]. The subtle point of choosing a state space partitioning for a "generic case" is discussed in ref. [15, 22].

**Remark 14.4** 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 [9, 3] you will see the right hand side of (14.33) defined as $\{B, A\}$ so that (14.34) is $\frac{dQ}{dt} = \{Q, H\}$. Other equally reputable texts [18] employ the convention used here. Landau and Lifshitz [4] 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 14.5** "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)

## Exercises

14.1. **Integrating over Dirac delta functions.** Let us verify a few of the properties of the delta function and check (14.9), as well as the formulas (14.7) and (14.8) to be used later.

(a) If $f : \mathbb{R}^d \to \mathbb{R}^d$, show that

$$\int_{\mathbb{R}^d} dx\, \delta(f(x)) = \sum_{x \in f^{-1}(0)} \frac{1}{|\det \partial_x f|} .$$

(b) The delta function can be approximated by a se-

quence of Gaussians

$$\int dx\,\delta(x)f(x) = \lim_{\sigma\to 0}\int dx\,\frac{e^{-\frac{x^2}{2\sigma}}}{\sqrt{2\pi\sigma}}f(x)\,.$$

Use this approximation to see whether the formal expression

$$\int_{\mathbb{R}} dx\,\delta(x^2)$$

makes sense.

14.2. **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

$$\int_{\mathbb{R}} dx\,\delta'(y) \quad , \qquad \text{where } y = f(x) - x \tag{14.41}$$

$$\int dx\,\delta^{(2)}(y) = \sum_{\{x:y(x)=0\}}\frac{1}{|y'|}\left\{3\frac{(y'')^2}{(y')^4} - \frac{y'''}{(y')^3}\right\} \tag{14.42}$$

$$\int dx\,b(x)\delta^{(2)}(y) = \sum_{\{x:y(x)=0\}}\frac{1}{|y'|}\left\{\frac{b''}{(y')^2} - \frac{b'y''}{(y')^3} + b\left(3\frac{(y'')^2}{(y')^4} - \frac{y'''}{(y')^3}\right)\right\} \tag{14.43}$$

These formulas are useful for computing effects of weak noise on deterministic dynamics [5].

14.3. **$\mathcal{L}^t$ generates a semigroup.** Check that the Perron-Frobenius operator has the semigroup property,

$$\int_M dz\mathcal{L}^{t_2}(y,z)\,\mathcal{L}^{t_1}(z,x) = \mathcal{L}^{t_2+t_1}(y,x)\,, \qquad t_1, t_2 \geq 0\,. \tag{14.44}$$

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 $t_1, t_2 \in \mathbb{R}$.

14.4. **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

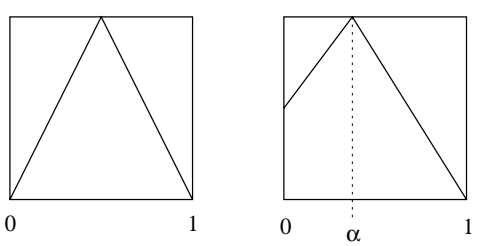
$$f(x) = a(1 - 2|x - 0.5|)$$

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).

14.5. **Invariant measure.** We will compute the invariant measure for two different piecewise linear maps.




(a) Verify the matrix $\mathcal{L}$ representation (15.19).

(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.

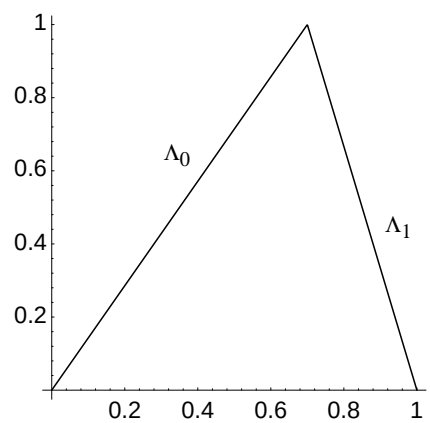
(d) For this map there is an infinite number of invariant measures, but only one of them will be found when one carries out a numerical simulation. Determine that measure, and explain why your choice is the natural measure for this map.

(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)

14.6. **Escape rate for a flow conserving map.** Adjust $\Lambda_0$, $\Lambda_1$ in (15.17) so that the gap between the intervals $\mathcal{M}_0$, $\mathcal{M}_1$ vanishes. Show that the escape rate equals zero in this situation.

14.7. **Eigenvalues of the Perron-Frobenius operator for the skew Ulam tent map.** Show that for the skew Ulam tent map




$$f(x) = \begin{cases} f_0(x) = \Lambda_0 x\,, & x \in \mathcal{M}_0 = [0, 1/\Lambda_0) \\ f_1(x) = \frac{\Lambda_0}{\Lambda_0 - 1}(1-x)\,, & x \in \mathcal{M}_1 = (1/\Lambda_0, 1] \end{cases}$$

the eigenvalues are available analytically, compute the first few.

14.8. **"Kissing disks"*** (continuation of exercises 8.1 and 8.2). 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 23.

14.9. **Invariant measure for the Gauss map.** Consider the Gauss map:

$$f(x) = \begin{cases} \frac{1}{x} - \left[\frac{1}{x}\right] & x \neq 0 \\ 0 & x = 0 \end{cases} \tag{14.46}$$

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?

14.10. **$\mathcal{A}$ as a generator of translations.** Verify that for a constant velocity field the evolution generator $\mathcal{A}$ in (14.30) is the generator of translations,

$$e^{tv\frac{\partial}{\partial x}} a(x) = a(x+tv)\,.$$

14.11. **Incompressible flows.** Show that (14.9) 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 14.8.

## References


[14.1] Ya.G. Sinai, *Topics in Ergodic Theory* (Princeton Univ. Press, Princeton, New Jersey 1994).

[14.2] J.E. Marsden and T.J.R. Hughes, *Mathematical Foundations of Elasticity* (Prentice-Hall, Englewood Cliffs, New Jersey 1983)

[14.3] H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, 1980).

[14.4] L.D. Landau and E.M. Lifshitz, *Mechanics* (Pergamon, London, 1959).

[14.5] 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 (1998); arXiv:chao-dyn/9807034.

[14.6] P. Cvitanović, C.P. Dettmann, R. Mainieri and G. Vattay, *Trace formulas for stochastic evolution operators: Smooth conjugation method*, *Nonlinearity* **12**, 939 (1999); arXiv:chao-dyn/9811003.

[14.7] P. Cvitanović, C.P. Dettmann, G. Palla, N. Søndergård and G. Vattay, *Spectrum of stochastic evolution operators: Local matrix representation approach*, *Phys. Rev.* **E 60**, 3936 (1999); arXiv:chao-dyn/9904027.

[14.8] A. Lasota and M.C. Mackey, *Chaos, Fractals and Noise* (Springer, New York 1994).

[14.9] G. W. Bluman and S. Kumei, *Symmetries and Differential Equations* (Springer, New York 1989).

[14.10] L. Billings and E.M. Bolt, "Invariant densities for skew tent maps," *Chaos Solitons and Fractals* **12**, 365 (2001); see also www.mathstat.concordia.ca/pg/bilbollt.html.

[14.11] G.D. Birkhoff, *Collected Math. Papers*, Vol. **II** (Amer. Math. Soc., Providence R.I., 1950).

[14.12] S. M. Ulam, *A Collection of Mathematical Problems* (Interscience Publishers, New York, 1960).

[14.13] G. Froyland, *Commun. Math. Phys.* **189**, 237 (1997).

[14.14] G. Froyland, *Discrete and Continuous Dynamical Systems* **17**, 671 (2007).

[14.15] G. Froyland, *Nonlinearity* **12**, 79 (1999).

[14.16] G. Froyland, "Extracting dynamical behaviour via Markov models," in A. Mees (ed.) *Nonlinear dynamics and statistics: Proceedings Newton Institute, Cambridge 1998* (Birkhäuser, 2000); math-www.uni-paderborn.de/ froyland.

[14.17] M. Dellnitz, G. Froyland and S. Sertl, *Nonlinearity* **13**, 1171 (2000).

[14.18] M.C. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer, New York 1990).

# Chapter 15

# Averaging

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*

We discuss first the necessity of studying the averages of observables in chaotic dynamics, and then cast 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. 15.3 we try out the formalism on the first quantitative diagnosis that a system's got chaos, Lyapunov exponents.

## 15.1 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. Hence for chaotic dynamics one cannot follow individual trajectories for a long time; what is attainable is a description of the geometry of the set of possible outcomes, and evaluation of long time averages. Examples of such averages are transport coefficients for chaotic dynamical flows, such as escape rate, mean drift and diffusion rate; 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 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.

### 15.1.1 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. It 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 length of this vector, or perhaps a temperature measured in an experiment at instant $\tau$ are examples of scalar observables. We define the *integrated observable* $A^t$ as the time integral of the observable $a$ evaluated along the trajectory of the initial point $x_0$,

$$A^t(x_0) = \int_0^t d\tau\, a(f^\tau(x_0))\,. \tag{15.1}$$

If the dynamics is given by an iterated mapping and the time is discrete, $t \to n$, the integrated observable is given by

$$A^n(x_0) = \sum_{k=0}^{n-1} a(f^k(x_0)) \tag{15.2}$$

(we suppress possible vectorial indices for the time being).

**Example 15.1** ***Integrated observables.*** *If the observable is the velocity, $a_i(x) = v_i(x)$, its time integral $A_i^t(x_0)$ is the trajectory $A_i^t(x_0) = x_i(t)$.*

*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:*

$$A^t(x_0) = \int_0^t d\tau\, \dot{\mathbf{q}}(\tau)\cdot\mathbf{p}(\tau)\,. \tag{15.3}$$

The *time average* of the observable along a trajectory is defined by

$$\overline{a(x_0)} = \lim_{t\to\infty}\frac{1}{t}A^t(x_0)\,. \tag{15.4}$$

If $a$ does not behave too wildly as a function of time – for example, if $a(x)$ is the Chicago temperature, bounded between $-80^oF$ and $+130^oF$ for all times – $A^t(x_0)$ is expected to grow not faster than $t$, and the limit (15.4) exists. For an example of a time average - the Lyapunov exponent - see sect. 15.3.

The time average depends on the trajectory, but not on the initial point on that trajectory: 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 $t \to \infty$ limit:

$$\begin{aligned}
\overline{a(f^T(x_0))} &= \lim_{t\to\infty}\frac{1}{t}\int_T^{t+T} d\tau\, a(f^\tau(x_0)) \\
&= \overline{a(x_0)} - \lim_{t\to\infty}\frac{1}{t}\left(\int_0^T d\tau\, a(f^\tau(x_0)) - \int_t^{t+T} d\tau\, a(f^\tau(x_0))\right) \\
&= \overline{a(x_0)}\,.
\end{aligned}$$

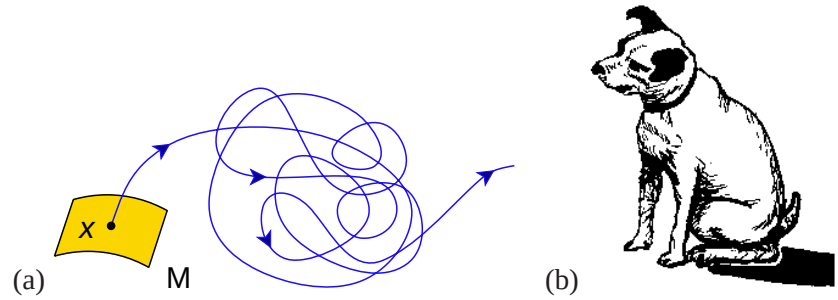



**Figure 15.1:** (a) A typical chaotic trajectory explores the phase space with the long time visitation frequency building up the natural measure $\rho_0(x)$. (b) time average evaluated along an atypical trajectory such as a periodic orbit fails to explore the entire accessible state space. (A. Johansen)

The integrated observable $A^t(x_0)$ and the time average $\overline{a(x_0)}$ take a particularly simple form when evaluated on a periodic orbit. Define

[exercise 4.6]

$$\begin{aligned} \text{flows:} \quad A_p &= a_p T_p = \int_0^{T_p} d\tau a\left(f^\tau(x_0)\right) , \qquad x_0 \in p \\ \text{maps:} \quad &= a_p n_p = \sum_{i=0}^{n_p-1} a\left(f^i(x_0)\right) , \end{aligned} \tag{15.5}$$

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. $A_p$ 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 $x_0 \in p$. (If the observable $a$ is not a scalar but a vector or matrix we might have to be more careful in defining an average which is independent of the starting point on the cycle). If the trajectory retraces itself $r$ times, we just obtain $A_p$ repeated $r$ times. Evaluation of the asymptotic time average (15.4) requires therefore only a single traversal of the cycle:

$$a_p = A_p/T_p \, . \tag{15.6}$$

However, $\overline{a(x_0)}$ is in general a wild function of $x_0$; for a hyperbolic system ergodic with respect to a smooth measure, it takes the same value $\langle a \rangle$ for almost all initial $x_0$, but a different value (15.6) on any periodic orbit, i.e., on a dense set of points (figure 15.1 (b)). For example, for an open system such as the Sinai gas of sect. 24.1 (an infinite 2-dimensional periodic array of scattering disks) the phase space 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 $D \approx x(t)^2/t$, (15.4) 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.

[chapter 24]

For chaotic dynamical systems, this paradox is resolved by robust averaging, i.e., averaging also over the initial $x$, and worrying about the measure of the "pathological" trajectories.

### 15.1.2 Space averages

The *space average* of a quantity $a$ that may depend on the point $x$ of state space $\mathcal{M}$ and on the time $t$ is given by the $d$-dimensional integral over the $d$ coordinates

of the dynamical system:

$$\begin{aligned}\langle a\rangle(t) &= \frac{1}{|\mathcal{M}|}\int_{\mathcal{M}} dx\, a(f^t(x))\\ |\mathcal{M}| &= \int_{\mathcal{M}} dx \;= \text{volume of } \mathcal{M}\,.\end{aligned} \tag{15.7}$$

The space $\mathcal{M}$ is assumed to have finite dimension and volume (open systems like the 3-disk game of pinball are discussed in sect. 15.1.3).

What is it we *really* do in experiments? We cannot measure the time average (15.4), as there is no way to prepare a single initial condition with infinite precision. The best we can do is to prepare some initial density $\rho(x)$ perhaps concentrated on some small (but always finite) neighborhood $\rho(x) = \rho(x, 0)$, so one should abandon the uniform space average (15.7), and consider instead

$$\langle a\rangle_\rho(t) = \frac{1}{|\mathcal{M}|}\int_{\mathcal{M}} dx\, \rho(x) a(f^t(x))\,. \tag{15.8}$$

We do not bother to lug the initial $\rho(x)$ around, as for the ergodic and mixing systems that we shall consider here *any* smooth initial density will tend to the asymptotic natural measure $t \to \infty$ limit $\rho(x, t) \to \rho_0(x)$, so we can just as well take the initial $\rho(x) = \text{const}$. The worst we can do is to start out with $\rho(x) = \text{const.}$, as in (15.7); so let us take this case and define the *expectation value* $\langle a\rangle$ of an observable $a$ to be the asymptotic time and space average over the state space $\mathcal{M}$

$$\langle a\rangle = \lim_{t\to\infty} \frac{1}{|\mathcal{M}|}\int_{\mathcal{M}} dx\, \frac{1}{t}\int_0^t d\tau\, a(f^\tau(x))\,. \tag{15.9}$$

We use the same $\langle\cdots\rangle$ notation as for the space average (15.7), 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 does not, it is the expectation value $\langle a\rangle$.

The expectation value is a space average of time averages, with every $x \in \mathcal{M}$ used as a starting point of a time average. The advantage of averaging over space is that it smears over the starting points which were problematic for the time average (like the 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

$$\left\langle e^{\beta\cdot A^t}\right\rangle = \frac{1}{|\mathcal{M}|}\int_{\mathcal{M}} dx\, e^{\beta\cdot A^t(x)}. \tag{15.10}$$

In the present context $\beta$ is an auxiliary variable of no particular physical significance. In most applications $\beta$ is a scalar, but if the observable is a $d$-dimensional

vector $a_i(x) \in \mathbb{R}^d$, then $\beta$ is a conjugate vector $\beta \in \mathbb{R}^d$; if the observable is a $d \times d$ tensor, $\beta$ is also a rank-2 tensor, and so on. Here we will mostly limit the considerations to scalar values of $\beta$.

If the limit $\overline{a(x_0)}$ for the time average (15.4) exists for "almost all" initial $x_0$ and the system is ergodic and mixing (in the sense of sect. 1.3.1), we expect the time average along almost all trajectories to tend to the same value $\overline{a}$, and the integrated observable $A^t$ to tend to $t\overline{a}$. The space average (15.10) is an integral over exponentials, and such integral also grows exponentially with time. So as $t \to \infty$ we would expect the space average of $\langle \exp(\beta \cdot A^t) \rangle$ itself to grow exponentially with time

$$\left\langle e^{\beta \cdot A^t} \right\rangle \propto e^{ts(\beta)} ,$$

and its rate of growth to go to a limit

$$s(\beta) = \lim_{t\to\infty} \frac{1}{t} \ln \left\langle e^{\beta \cdot A^t} \right\rangle . \tag{15.11}$$

Now we understand one reason for why it is smarter to compute $\langle \exp(\beta \cdot A^t) \rangle$ rather than $\langle a \rangle$: the expectation value of the observable (15.9) and the moments of the integrated observable (15.1) can be computed by evaluating the derivatives of $s(\beta)$

$$\begin{aligned} \left.\frac{\partial s}{\partial \beta}\right|_{\beta=0} &= \lim_{t\to\infty} \frac{1}{t} \left\langle A^t \right\rangle = \langle a \rangle , \\ \left.\frac{\partial^2 s}{\partial \beta^2}\right|_{\beta=0} &= \lim_{t\to\infty} \frac{1}{t} \left( \left\langle A^t A^t \right\rangle - \left\langle A^t \right\rangle \left\langle A^t \right\rangle \right) \\ &= \lim_{t\to\infty} \frac{1}{t} \left\langle (A^t - t\langle a \rangle)^2 \right\rangle , \end{aligned} \tag{15.12}$$

and so forth. We have written out the formulas for a scalar observable; the vector case is worked out in the exercise 15.2. 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 15.2]

Suppose we could evaluate $s(\beta)$ and its derivatives. What are such formulas good for? A typical application is to the problem of describing a particle scattering elastically off a 2-dimensional triangular array of disks. If the disks 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 given by $\left\langle x(t)^2 \right\rangle \approx 4Dt$. 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 $t \to \infty$ 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 $\beta = 0$,

$$D = \lim_{t\to\infty} \frac{1}{2dt} \left\langle x(t)^2 \right\rangle = \frac{1}{2d} \sum_{i=1}^{d} \left.\frac{\partial^2 s}{\partial \beta_i^2}\right|_{\beta=0} , \tag{15.13}$$

[section 24.1]

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 explicit closed form expression for $D$.

fast track:
sect. 15.2, p. 261

### 15.1.3 Averaging in open systems

If the $\mathcal{M}$ is a compact region or set of regions to which the dynamics is confined for all times, (15.9) is a sensible definition of the expectation value. However, if the trajectories can exit $\mathcal{M}$ without ever returning,

$$\int_{\mathcal{M}} dy\, \delta(y - f^t(x_0)) = 0 \qquad \text{for } t > t_{exit}\,, \qquad x_0 \in \mathcal{M}\,,$$

we might be in trouble. In particular, for a repeller the trajectory $f^t(x_0)$ will eventually leave the region $\mathcal{M}$, unless the initial point $x_0$ is on the repeller, so the identity

$$\int_{\mathcal{M}} dy\, \delta(y - f^t(x_0)) = 1\,, \quad t > 0\,, \qquad \text{iff } x_0 \in \text{non–wandering set} \tag{15.14}$$

might apply only to a fractal subset of initial points a set of zero Lebesgue measure. 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.

Note 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 the state space containing all trajectories which start out within the state space region $\mathcal{M}$ and recur within that region at the time $t$

$$|\mathcal{M}(t)| = \int_{\mathcal{M}} dx dy\, \delta\Big(y - f^t(x)\Big) \;\sim\; |\mathcal{M}| e^{-\gamma t} \tag{15.15}$$

is expected to decrease exponentially, with the escape rate $\gamma$. 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 figure 1.1 is gone for good.

[section 1.4.3]

[section 20.1]

The non–wandering set can be very difficult object to describe; but for any finite time we can construct a normalized measure from the finite-time covering volume (15.15), by redefining the space average (15.10) as

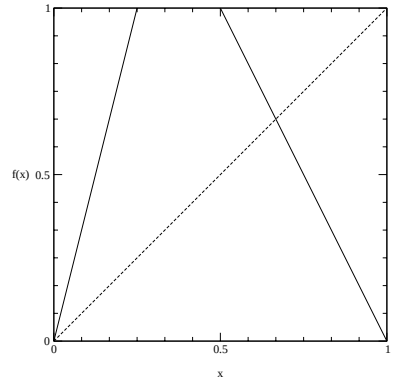



**Figure 15.2:** A piecewise-linear repeller (15.17): All trajectories that land in the gap between the $f_0$ and $f_1$ branches escape ($\Lambda_0 = 4$, $\Lambda_1 = -2$).

$$\left\langle e^{\beta\cdot A^t}\right\rangle = \int_{\mathcal{M}} dx\, \frac{1}{|\mathcal{M}(t)|} e^{\beta\cdot A^t(x)} \sim \frac{1}{|\mathcal{M}|}\int_{\mathcal{M}} dx\, e^{\beta\cdot A^t(x)+\gamma t}\,. \tag{15.16}$$

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 $\gamma$ we can replenish the density lost to escaping trajectories, by pumping in $e^{\gamma t}$ in such a way that the overall measure is correctly normalized at all times, $\langle 1\rangle = 1$.

**Example 15.2** ***A piecewise-linear repeller:*** *(continuation of example 14.1) 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-d map* $f(x)$ *of figure 15.2, a piecewise-linear 2–branch repeller with slopes* $\Lambda_0 > 1$ *and* $\Lambda_1 < -1$ *:*

$$f(x) = \begin{cases} f_0 = \Lambda_0 x & \text{if} \quad x \in \mathcal{M}_0 = [0, 1/\Lambda_0] \\ f_1 = \Lambda_1(x-1) & \text{if} \quad x \in \mathcal{M}_1 = [1+1/\Lambda_1, 1] \end{cases}\,. \tag{15.17}$$

*Both* $f(\mathcal{M}_0)$ *and* $f(\mathcal{M}_1)$ *map onto the entire unit interval* $\mathcal{M} = [0,1]$*. Assume a piecewise constant density*

$$\rho(x) = \begin{cases} \rho_0 & \text{if } x \in \mathcal{M}_0 \\ \rho_1 & \text{if } x \in \mathcal{M}_1 \end{cases}\,. \tag{15.18}$$

*There is no need to define* $\rho(x)$ *in the gap between* $\mathcal{M}_0$ *and* $\mathcal{M}_1$*, 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 (14.9), the Perron-Frobenius operator acts on this piecewise constant function as a [2×2] "transfer" matrix with matrix elements*

[exercise 14.1]

[exercise 14.5]

$$\begin{pmatrix}\rho_0\\ \rho_1\end{pmatrix} \to \mathcal{L}\rho = \begin{pmatrix} \frac{1}{|\Lambda_0|} & \frac{1}{|\Lambda_1|} \\ \frac{1}{|\Lambda_0|} & \frac{1}{|\Lambda_1|}\end{pmatrix}\begin{pmatrix}\rho_0\\ \rho_1\end{pmatrix}, \tag{15.19}$$

*stretching both* $\rho_0$ *and* $\rho_1$ *over the whole unit interval* $\Lambda$*, 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* $e^{s_0} = 1/|\Lambda_0| + 1/|\Lambda_1|$*, with constant density eigenvector* $\rho_0 = \rho_1$*. The quantities* $1/|\Lambda_0|$*,* $1/|\Lambda_1|$ *are, respectively, the sizes of the* $|\mathcal{M}_0|$*,* $|\mathcal{M}_1|$

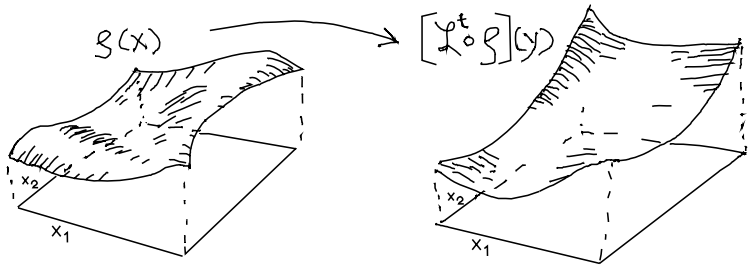



**Figure 15.3:** Space averaging pieces together the time average computed along the $t \to \infty$ trajectory of figure 15.1 by a space average over infinitely many short $t$ trajectory segments starting at all initial points at once. (A. Johansen)

*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 $\mathcal{L}$:*

$$\gamma = -s_0 = -\ln(1/|\Lambda_0| + 1/|\Lambda_1|)\,. \tag{15.20}$$

*Voila! Here is the rationale for introducing operators – in one time step we have solved the problem of evaluating escape rates at infinite time. 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 $\rho$ to the space of piecewise constant functions. The example gives a flavor of the enterprise upon which we are about to embark in this book, but the full story is much subtler: in general, there will exist no such finite-dimensional representation for the Perron-Frobenius operator.*

We now turn to the problem of evaluating $\left\langle e^{\beta\cdot A^t}\right\rangle$.

## 15.2 Evolution operators

The above simple shift of focus, from studying $\langle a\rangle$ to studying $\langle \exp(\beta\cdot A^t)\rangle$ is the key to all that follows. Make the dependence on the flow explicit by rewriting this quantity as

$$\left\langle e^{\beta\cdot A^t}\right\rangle = \frac{1}{|\mathcal{M}|}\int_{\mathcal{M}} dx \int_{\mathcal{M}} dy\, \delta\Big(y - f^t(x)\Big)\, e^{\beta\cdot A^t(x)}\,. \tag{15.21}$$

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

$$1 = \int_{\mathcal{M}} dy\, \delta\Big(y - f^t(x)\Big)\,, \tag{15.22}$$

into (15.10) 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 the study of the evolution of density of *the totality* of initial conditions. Instead of trying to extract a temporal average from an arbitrarily long trajectory which explores the phase 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 (14.27) of infinitesimal transformations of densities) *infinitesimally* short time evolution can suffice to determine the spectrum and eigenvalues of $\mathcal{L}^t$.

We shall refer to the kernel of the operation (15.21) as $\mathcal{L}^t(y,x)$.

$$\mathcal{L}^t(y,x) \;=\; \delta\Big(y - f^t(x)\Big)\, e^{\beta\cdot A^t(x)}\,. \tag{15.23}$$

The evolution operator acts on scalar functions $\phi(x)$ as

$$(y) = \int_{\mathcal{M}} dx\, \delta\Big(y - f^t(x)\Big)\, e^{\beta\cdot A^t(x)}\phi(x)\,. \tag{15.24}$$

In terms of the evolution operator, the space average of the generating function (15.21) is given by

$$\Big\langle e^{\beta\cdot A^t}\Big\rangle = \langle\rangle\ ,$$

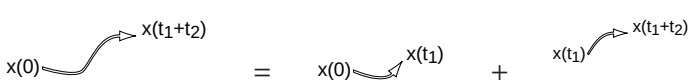
and, if the spectrum of the linear operator $\mathcal{L}^t$ can be described, by (15.11) this limit

$$s(\beta) = \lim_{t\to\infty}\frac{1}{t}\ln\Big\langle\mathcal{L}^t\Big\rangle\ .$$

yields the leading eigenvalue of $s_0(\beta)$, and, through it, all desired expectation values (15.12).

The evolution operator is different for different observables, as its definition depends on the choice of the integrated observable $A^t$ in the exponential. Its job is deliver to us 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




$$\begin{aligned} A^{t_1+t_2}(x_0) &= \int_0^{t_1} d\tau &&+ \int_{t_1}^{t_1+t_2} d\tau \\ &= A^{t_1}(x_0) &&+ A^{t_2}(f^{t_1}(x_0))\,. \end{aligned}$$

[exercise 14.3]

If $A^t(x)$ is additive along the trajectory, the evolution operator generates a semigroup

[section 14.5]

$$\mathcal{L}^{t_1+t_2}(y,x) = \int_{\mathcal{M}} dz\, \mathcal{L}^{t_2}(y,z)\mathcal{L}^{t_1}(z,x)\,, \tag{15.25}$$

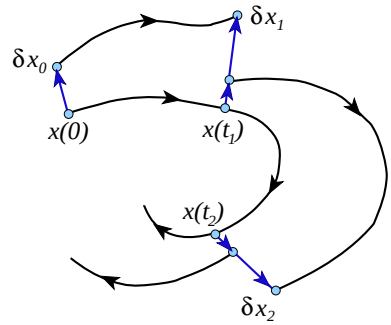



**Figure 15.4:** A long-time numerical calculation of the leading Lyapunov exponent requires rescaling the distance in order to keep the nearby trajectory separation within the linearized flow range.

as is easily checked by substitution

$$\mathcal{L}^{t_2}\mathcal{L}^{t_1}a(y) = \int_{\mathcal{M}} dx\, \delta(y - f^{t_2}(x))e^{\beta\cdot A^{t_2}(x)}(\mathcal{L}^{t_1}a)(x) = \mathcal{L}^{t_1+t_2}a(y)\,.$$

This semigroup property is the main reason why (15.21) is preferable to (15.9) as a starting point for evaluation of dynamical averages: it recasts averaging in form of operators multiplicative along the flow.

## 15.3 Lyapunov exponents

(J. Mathiesen and P. Cvitanović)

Let us apply the newly acquired tools to the fundamental diagnostics in this subject: Is a given system "chaotic"? And if so, how chaotic? If all points in a neighborhood of a trajectory converge toward the same trajectory, the attractor is a fixed point or a limit cycle. However, if the attractor is strange, any two trajectories [example 2.3] [section 1.3.1]

$$x(t) = f^t(x_0) \quad \text{and} \quad x(t) + \delta x(t) = f^t(x_0 + \delta x_0) \tag{15.26}$$

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. This *sensitivity to initial conditions* can be quantified as

$$|\delta x(t)| \approx e^{\lambda t}|\delta x_0| \tag{15.27}$$

where $\lambda$, the mean rate of separation of trajectories of the system, is called the *Lyapunov exponent*.

### 15.3.1 Lyapunov exponent as a time average

We can start out with a small $\delta x$ and try to estimate $\lambda$ from (15.27), but now that we have quantified the notion of linear stability in chapter4 and defined the dynamical

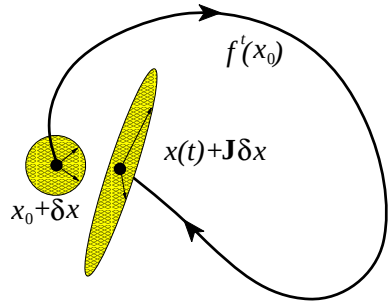



**Figure 15.5:** The symmetric matrix $\left(J^t\right)^T J^t$ maps a swarm of initial points in an infinitesimal spherical neighborhood of $x_0$ into a cigar-shaped neighborhood finite time $t$ later, with semiaxes determined by the local stretching/shrinking $|\Lambda_1|$, but local individual trajectory rotations by the complex phase of $J^t$ ignored.

time averages in sect. 15.1.1, 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 study of experimental time series this might be the only option, but if we have the equations of motion, a better way is to measure the growth rate of vectors transverse to a given orbit.

The mean growth rate of the distance $|\delta x(t)|/|\delta x_0|$ between neighboring trajectories (15.27) is given by the *Lyapunov exponent*

$$\lambda = \lim_{t\to\infty} \frac{1}{t} \ln |\delta x(t)|/|\delta x_0| \tag{15.28}$$

(For notational brevity we shall often suppress the dependence of quantities such as $\lambda = \lambda(x_0)$, $\delta x(t) = \delta x(x_0, t)$ on the initial point $x_0$ and the time $t$). One can take (15.28) as is, take a small initial separation $\delta x_0$, track distance between two nearby trajectories until $|\delta x(t_1)|$ gets significantly bigger, 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, with the leading Lyapunov exponent given by

$$\lambda = \lim_{t\to\infty} \frac{1}{t} \sum_i t_i \lambda_i \,. \tag{15.29}$$

However, we can do better. Given the equations of motion and baring numerical problems (such as evaluating the fundamental matrix (4.43) for high-dimensional flows), for infinitesimal $\delta x$ we know the $\delta x_i(t)/\delta x_j(0)$ ratio exactly, as this is by definition the fundamental matrix (4.43)

$$\lim_{\delta x\to 0} \frac{\delta x_i(t)}{\delta x_j(0)} = \frac{\partial x_i(t)}{\partial x_j(0)} = J^t_{ij}(x_0)\,,$$

so the leading Lyapunov exponent can be computed from the linear approximation (4.28)

$$\lambda = \lim_{t\to\infty} \frac{1}{t} \ln \frac{\left|J^t(x_0)\delta x_0\right|}{|\delta x_0|} = \lim_{t\to\infty} \frac{1}{2t} \ln \left|\hat{n}^T (J^t)^T J^t \hat{n}\right| \,. \tag{15.30}$$

In this formula the scale of the initial separation drops out, only its orientation given by the initial orientation unit vector $\hat{n} = \delta x/|\delta x|$ matters. The eigenvalues of $J$ are either real or come in complex conjugate pairs. As $J$ is in general

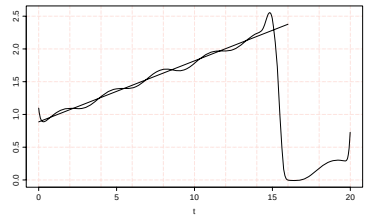



**Figure 15.6:** A numerical estimate of the leading Lyapunov exponent for the Rössler flow (2.17) from the dominant expanding eigenvalue formula (15.30). The leading Lyapunov exponent $\lambda \approx 0.09$ is positive, so numerics supports the hypothesis that the Rössler attractor is strange. (J. Mathiesen)

not symmetric and not diagonalizable, it is more convenient to work with the symmetric and diagonalizable matrix $\mathbf{J} = (J^t)^T J^t$, with real positive eigenvalues $\{|\Lambda_1|^2 \geq \ldots \geq |\Lambda_d|^2\}$, and a complete orthonormal set of eigenvectors of $\{u_1, \ldots, u_d\}$. Expanding the initial orientation $\hat{n} = \sum(\hat{n} \cdot u_i)u_i$ in the $\mathbf{J}u_i = u_i$ eigenbasis, we have

$$\hat{n}^T \mathbf{J}\hat{n} = \sum_{i=1}^{d} (\hat{n} \cdot u_i)^2 |\Lambda_i|^2 = (\hat{n} \cdot u_1)^2 e^{2\lambda_1 t} \left(1 + O(e^{-2(\lambda_1 - \lambda_2)t})\right) , \tag{15.31}$$

where $t\lambda_i = \ln|\Lambda_i(x_0, t)|$, with exponents ordered by $\lambda_1 > \lambda_2 \geq \lambda_3 \cdots$. For long times the largest Lyapunov exponent dominates exponentially (15.30), provided the orientation $\hat{n}$ of the initial separation was not chosen perpendicular to the dominant expanding eigendirection $u_1$. The Lyapunov exponent is the time average

$$\begin{aligned} \overline{\lambda(x_0)} &= \lim_{t\to\infty} \frac{1}{t} \left\{ \ln|\hat{n} \cdot u_1| + \ln|\Lambda_1(x_0, t)| + O(e^{-2(\lambda_1 - \lambda_2)t}) \right\} \\ &= \lim_{t\to\infty} \frac{1}{t} \ln|\Lambda_1(x_0, t)| , \end{aligned} \tag{15.32}$$

where $\Lambda_1(x_0, t)$ is the leading eigenvalue of $J^t(x_0)$. By choosing the initial displacement such that $\hat{n}$ is normal to the first ($i$-1) eigendirections we can define not only the leading, but all Lyapunov exponents as well:

$$\overline{\lambda_i(x_0)} = \lim_{t\to\infty} \frac{1}{t} \ln|\Lambda_i(x_0, t)| , \qquad i = 1, 2, \cdots, d . \tag{15.33}$$

The leading Lyapunov exponent now follows from the fundamental matrix by numerical integration of (4.9).

The equations can be integrated accurately for a finite time, hence the infinite time limit of (15.30) can be only estimated from plots of $\frac{1}{2} \ln|\hat{n}^T \mathbf{J}\hat{n}|$ as function of time, such as the figure 15.6 for the Rössler flow (2.17).

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 level is caused by a close passage to a folding point of the attractor, an illustration of why numerical evaluation of the Lyapunov exponents, and proving the very existence of a strange attractor is a very difficult problem. The approximately monotone part of the curve can be used (at your own peril) 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 (15.32) directly. First of all, the state space is dense with atypical trajectories; for example, if $x_0$ happened to lie on a periodic orbit $p$, $\overline{\lambda}$ would be simply $\ln|\Lambda_p|/T_p$, 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 $\ln|\Lambda(x_0,t)|/t$ should be converging to anything in particular. In a Hamiltonian system with coexisting elliptic islands and chaotic regions, a chaotic trajectory gets every so often captured in the neighborhood of an elliptic island and can stay there for arbitrarily long time; as there the orbit is nearly stable, during such episode $\ln|\Lambda(x_0,t)|/t$ can dip arbitrarily close to $0^+$. For state space volume non-preserving flows the trajectory can traverse locally contracting regions, and $\ln|\Lambda(x_0,t)|/t$ can occasionally go negative; even worse, one never knows whether the asymptotic attractor is periodic or "strange," so any finite estimate of $\overline{\lambda}$ might be dead wrong.

[exercise 15.1]

### 15.3.2 Evolution operator evaluation of Lyapunov exponents

A cure to these problems was offered in sect. 15.2. We shall now replace time averaging along a single trajectory by action of a multiplicative evolution operator on the entire state space, and extract the Lyapunov exponent from its leading eigenvalue. 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 non-wandering set, will characterize the chaotic transient; this is actually what any experiment would measure, as even very small amount of external noise will suffice to destabilize a long stable cycle with a minute immediate basin of attraction.

Due to the chain rule (4.51) for the derivative of an iterated map, the stability of a 1-$d$ mapping is multiplicative along the flow, so the integral (15.1) of the observable $a(x) = \ln|f'(x)|$, the local trajectory divergence rate, evaluated along the trajectory of $x_0$ is additive:

$$A^n(x_0) = \ln\left|f^{n\prime}(x_0)\right| = \sum_{k=0}^{n-1} \ln\left|f'(x_k)\right| \,. \tag{15.34}$$

The Lyapunov exponent is then the expectation value (15.9) given by a spatial integral (15.8) weighted by the natural measure

$$\lambda = \langle \ln|f'(x)| \rangle = \int_{\mathcal{M}} dx\, \rho_0(x) \ln|f'(x)| \,. \tag{15.35}$$

The associated (discrete time) evolution operator (15.23) is

$$\mathcal{L}(y,x) \;=\; \delta(y - f(x))\; e^{\beta \ln|f'(x)|} \,. \tag{15.36}$$

Here we have restricted our considerations to $1-d$ maps, as for higher-dimensional flows only the fundamental matrices are multiplicative, not the individual eigenvalues. Construction of the evolution operator for evaluation of the Lyapunov spectra in the general case requires more cleverness than warranted at this stage in the narrative: an extension of the evolution equations to a flow in the tangent space.

All that remains is to determine the value of the Lyapunov exponent

$$\lambda = \langle \ln |f'(x)| \rangle = \left. \frac{\partial s(\beta)}{\partial \beta} \right|_{\beta=1} = s'(1) \tag{15.37}$$

from (15.12), the derivative of the leading eigenvalue $s_0(\beta)$ of the evolution operator (15.36). [example 18.1]

The only question is: how?

## Résumé

The expectation value $\langle a \rangle$ of an observable $a(x)$ measured $A^t(x) = \int_0^t d\tau a(x(\tau))$ and averaged along the flow $x \to f^t(x)$ 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 corresponding evolution operator $\mathcal{L}^t$.

Instead of using the Perron-Frobenius operator (14.10) whose leading eigenfunction, the natural measure, once computed, yields expectation value (14.20) of any observable $a(x)$, we construct a specific, hand-tailored evolution operator $\mathcal{L}$ for each and every observable. However, by time we arrive to chapter 18, 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 18]

The next question is: how do we evaluate the eigenvalues of $\mathcal{L}$? We saw in example 15.2, in the case of piecewise-linear dynamical systems, that these operators reduce to finite matrices, but for generic smooth flows, they are infinite-dimensional linear operators, and finding smart ways of computing their eigenvalues requires some thought. In chapter 10 we undertook the first step, and replaced the *ad hoc* partitioning (14.14) by the intrinsic, topologically invariant partitioning. In chapter 13 we applied this information to our first application of the evolution operator formalism, evaluation of the topological entropy, the growth rate of the number of topologically distinct orbits. This small victory will be refashioned in chapters 16 and 17 into a systematic method for computing eigenvalues of evolution operators in terms of periodic orbits.

## Commentary

**Remark 15.1** "Pressure." The quantity $\langle \exp(\beta \cdot A^t) \rangle$ is called a "partition function" by Ruelle [1]. Mathematicians decorate it with considerably more Greek and Gothic letters than is the case in this treatise. Ruelle [1] and Bowen [2] had given name "pressure" $P(a)$ to $s(\beta)$ (where $a$ is the observable introduced here in sect. 15.1.1), defined by the "large system" limit (15.11). As we shall apply the theory also to computation of the physical gas pressure exerted on the walls of a container by a bouncing particle, we prefer to refer to $s(\beta)$ as simply the leading eigenvalue of the evolution operator introduced in sect. 14.5. The "convexity" properties such as $P(a) \leq P(|a|)$ will be pretty obvious consequence of the definition (15.11). In the case that $\mathcal{L}$ is the Perron-Frobenius operator (14.10), the eigenvalues $\{s_0(\beta), s_1(\beta), \cdots\}$ are called the *Ruelle-Pollicott resonances* [3, 4, 5], with the leading one, $s(\beta) = s_0(\beta)$ being the one of main physical interest. In order to aid the reader in digesting the mathematics literature, we shall try to point out the notational correspondences whenever appropriate. The rigorous formalism is replete with lims, sups, infs, $\Omega$-sets which are not really essential to understanding of the theory, and are avoided in this presentation.

**Remark 15.2** Microcanonical ensemble. In statistical mechanics the space average (15.7) performed over the Hamiltonian system constant energy surface invariant measure $\rho(x)dx = dqdp\,\delta(H(q,p) - E)$ of volume $\omega(E) = \int_{\mathcal{M}} dqdp\,\delta(H(q,p) - E)$

$$\langle a(t) \rangle = \frac{1}{\omega(E)} \int_{\mathcal{M}} dqdp\,\delta(H(q,p) - E) a(q,p,t) \tag{15.38}$$

is called the *microcanonical ensemble average*.

**Remark 15.3** Lyapunov exponents. The Multiplicative Ergodic Theorem of Oseledec [6] states that the limits (15.30–15.33) exist for almost all points $x_0$ and all tangent vectors $\hat{n}$. There are at most $d$ distinct values of $\lambda$ as we let $\hat{n}$ range over the tangent space. These are the Lyapunov exponents [8] $\lambda_i(x_0)$.

There is much literature on numerical computation of the Lyapunov exponents, see for example refs. [14, 15, 16].

**Remark 15.4** State space discretization. Ref. [17] discusses numerical discretizatons 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. [18].

## Exercises

15.1. **How unstable is the Hénon attractor?**

(a) Evaluate numerically the Lyapunov exponent $\lambda$ by iterating 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) Now check 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 trust now your result for the part (a) of this exercise?

15.2. **Expectation value of a vector observable.**

Check and extend the expectation value formulas (15.12) by evaluating the derivatives of $s(\beta)$ up to 4-th order for the space average $\langle \exp(\beta \cdot A^t) \rangle$ with $a_i$ a vector quantity:

(a)

$$\left.\frac{\partial s}{\partial \beta_i}\right|_{\beta=0} = \lim_{t\to\infty} \frac{1}{t} \left\langle A_i^t \right\rangle = \langle a_i \rangle \ , \quad (15.39)$$

(b)

$$\left.\frac{\partial^2 s}{\partial \beta_i \partial \beta_j}\right|_{\beta=0} = \lim_{t\to\infty} \frac{1}{t} \left( \left\langle A_i^t A_j^t \right\rangle - \left\langle A_i^t \right\rangle \left\langle A_j^t \right\rangle \right)$$

$$= \lim_{t\to\infty} \frac{1}{t} \left\langle (A_i^t - t\langle a_i \rangle)(A_j^t - t\langle a_j \rangle) \right\rangle \quad (15.40)$$

Note that the formalism is smart: it automatically yields the *variance* from the mean, rather than simply the 2nd moment $\left\langle a^2 \right\rangle$.

(c) compute the third derivative of $s(\beta)$.

(d) compute the fourth derivative assuming that the mean in (15.39) vanishes, $\langle a_i \rangle = 0$. The 4-th order moment formula

$$K(t) = \frac{\left\langle x^4(t) \right\rangle}{\langle x^2(t) \rangle^2} - 3 \quad (15.41)$$

that you have derived is known as *kurtosis*: it measures a deviation from what the 4-th order moment would be were the distribution a pure Gaussian (see (24.22) for a concrete example). If the observable is a vector, the kurtosis $K(t)$ is given by

$$\frac{\sum_{ij} \left[ \left\langle A_i A_i A_j A_j \right\rangle + 2 \left( \left\langle A_i A_j \right\rangle \left\langle A_j A_i \right\rangle - \langle A_i A_i \rangle \left\langle A_j \right.\right.\right.}{\left(\sum_i \langle A_i A_i \rangle\right)^2}$$

15.3. **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 $\gamma = .410\ldots$.

15.4. **Rössler attractor Lyapunov exponents.**

(a) Evaluate numerically the expanding Lyapunov exponent $\lambda_e$ of the Rössler attractor (2.17).

(b) Plot your own version of figure 15.6. 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 $\lambda_e$. The literature gives surprisingly inaccurate estimates - see whether you can do better.

(d) Estimate the contracting Lyapunov exponent $\lambda_c$. Even though it is much smaller than $\lambda_e$, a glance at the stability matrix (4.4) suggests that you can probably get it by integrating the infinitesimal volume along a long-time trajectory, as in (4.47).

## References


[15.1] D. Ruelle, *Bull. Amer. Math. Soc.* **78**, 988 (1972).

[15.2] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms, Springer Lect. Notes on Math.* **470** (1975).

[15.3] D. Ruelle, "The thermodynamical formalism for expanding maps," *J. Diff. Geo.* **25**, 117 (1987).

[15.4] M. Pollicott, "On the rate of mixing of Axiom A flows," *Invent. Math.* **81**, 413 (1985).

[15.5] D. Ruelle, *J. Diff. Geo.* **25**, 99 (1987).

[15.6] V. I. Oseledec, *Trans. Moscow Math. Soc.* **19**, 197 (1968).

[15.7] M. Pollicott, *Lectures on Ergodic Theory and Pesin Theory in Compact Manifolds*, (CUP, Cambridge 1993).

[15.8] A. M. Lyapunov, *General problem of stability of motion*, *Ann.Math. Studies* **17** (1949) (Princeton Univ. Press).

[15.9] Ya.B. Pesin, *Uspekhi Mat. Nauk* **32,** 55 (1977), [*Russian Math. Surveys* **32,** 55 (1977)]

[15.10] Ya.B. Pesin, *Dynamical systems with generalized hyperbolic attractors: hyperbolic, ergodic and topological properties*, *Ergodic Theory and Dynamical Systems*, **12**, 123 (1992).

[15.11] Ya.B. Pesin, *Func. Anal. Applic.* **8**, 263 (1974).

[15.12] A. Katok, *Lyapunov exponents, entropy and periodic orbits for diffeomorphisms*, *Publ. Math. IHES* **51**, 137 (1980).

[15.13] D. Bessis, G. Paladin, G. Turchetti and S. Vaienti, *Generalized Dimensions, Entropies and Lyapunov Exponents from the Pressure Function for Strange Sets*, *J. Stat. Phys.* **51**, 109 (1988).

[15.14] A. Wolf, J.B. Swift, et al., "Determining Lyapunov Exponents from a Time Series," *Physica D* **16**, 285 (1985).

[15.15] J.-P. Eckmann, S.O. Kamphorst, et al., "Lyapunov exponents from time series," *Phys. Rev. A* **34**, 4971 (1986).

[15.16] G. Benettin, L. Galgani, A. Giorgilli and J.-M. Strelcyn, "Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems: a method for computing all of them. Part 1:Theory," *Meccanica* **15**, 9 (1980); "Part 2: Numerical Application," *Meccanica* **15**, 21 (1980).

[15.17] 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," *Internat. J. Bifur. Chaos* **15**, 699 (2005); www.cds.caltech.edu/~koon/papers

[15.18] G. Froyland, "Computer-assisted bounds for the rate of decay of correlations," *Commun.Math.Phys.* **189**, 237 (1997); C. Liverani, "Rigorous numerical investigation of the statistical properties of piecewise expanding maps. A feasibility study," *Nonlinearity* **14**, 463 (2001).

# Chapter 16

# 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? We have given a quick sketch of this program in sects. 1.5 and 1.6; now we redo the same material in greater depth. In chapter 15 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 ↔ 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.

## 16.1 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. 16.2, for flows. The final formula (16.23) covers both cases.

To compute an expectation value using (15.21) we have to integrate over all the values of the kernel $\mathcal{L}^n(x,y)$. If $\mathcal{L}^n$ were a matrix we would be computing a

weighted sum of its eigenvalues which is dominated by the leading eigenvalue as $n \to \infty$. As the trace of $\mathcal{L}^n$ is also dominated by the leading eigenvalue as $t \to \infty$, we might just as well look at the trace

[exercise 13.2]

$$\mathrm{tr}\,\mathcal{L}^n = \int dx\, \mathcal{L}^n(x,x) = \int dx\, \delta(x - f^n(x))\, e^{\beta\cdot A^n(x)} \,. \tag{16.1}$$

By definition, the trace is the sum over eigenvalues,

$$\mathrm{tr}\,\mathcal{L}^n = \sum_{\alpha=0}^{\infty} e^{s_\alpha n} \,. \tag{16.2}$$

We find it convenient to write the eigenvalues as exponents $e^{s_\alpha}$ rather than as multipliers $\lambda_\alpha$, and we assume that spectrum of $\mathcal{L}$ is discrete, $s_0, s_1, s_2, \cdots$, ordered so that $\mathrm{Re}\ \mathrm{s}_\alpha \geq \mathrm{Re}\,\mathrm{s}_{\alpha+1}$.

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 21, and discuss how lack of hyperbolicity leads to continuous spectra in chapter 23.

### 16.1.1 Hyperbolicity assumption

We have learned in sect. 14.2 how to evaluate the delta-function integral (16.1).

According to (14.8) the trace (16.1) 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. 16.2, 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,

$$\begin{aligned} \mathrm{tr}\,{}_p\mathcal{L}^{n_p} &= \int_{\mathcal{M}_p} dx\, \delta(x - f^{n_p}(x)) \\ &= \frac{n_p}{\left|\det\left(\mathbf{1} - M_p\right)\right|} = n_p \prod_{i=1}^{d} \frac{1}{|1 - \Lambda_{p,i}|} \,. \end{aligned} \tag{16.3}$$

For the time being we set here and in (14.9) the observable $e^{\beta A_p} = 1$. Periodic orbit fundamental matrix $M_p$ is also known as the *monodromy matrix*, and its eigenvalues $\Lambda_{p,1}, \Lambda_{p,2}, \ldots, \Lambda_{p,d}$ as the Floquet multipliers.

We sort the eigenvalues $\Lambda_{p,1}, \Lambda_{p,2}, \ldots, \Lambda_{p,d}$ of the $p$-cycle $[d\times d]$ fundamental matrix $M_p$ into expanding, marginal and contracting sets $\{e, m, c\}$, as in (5.5).

As the integral (16.3) 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. 16.2 that the longitudinal $\Lambda_{p,d+1} = 1$ eigenvalue for flows can be eliminated by restricting the consideration to the transverse fundamental matrix $M_p$), and factorize the trace (16.3) into a product over the expanding and the contracting eigenvalues

$$\left|\det\left(\mathbf{1} - M_p\right)\right|^{-1} = \frac{1}{|\Lambda_p|}\prod_e \frac{1}{1 - 1/\Lambda_{p,e}}\prod_c \frac{1}{1 - \Lambda_{p,c}} \ , \tag{16.4}$$

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:

$$\begin{aligned} |\Lambda_{p,e}| &> e^{\lambda_e T_p} && \text{any } p\text{, any expanding } |\Lambda_{p,e}| > 1 \\ |\Lambda_{p,c}| &< e^{-\lambda_c T_p} && \text{any } p\text{, any contracting } |\Lambda_{p,c}| < 1 \ , \end{aligned} \tag{16.5}$$

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 23.

*Elliptic stability*, with 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

$$\frac{1}{(1 - e^{i\theta})(1 - e^{-i\theta})} = \frac{1}{2(1 - \cos\theta)} \tag{16.6}$$

does not make (14.9) diverge, if $\Lambda_m = e^{i2\pi p/r}$ is $r$th root of unity, $1/\left|\det\left(\mathbf{1} - M_p^r\right)\right|$ diverges. For a generic $\theta$ 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.

It follows from (16.4) that for long times, $t = rT_p \to \infty$, only the product of expanding eigenvalues matters, $\left|\det\left(\mathbf{1} - M_p^r\right)\right| \to |\Lambda_p|^r$. We shall use this fact to motivate the construction of dynamical zeta functions in sect. 17.3. However, for evaluation of the full spectrum the exact cycle weight (16.3) has to be kept.

### 16.1.2 A classical trace formula for maps

If the evolution is given by a discrete time mapping, and all periodic points have stability eigenvalues $|\Lambda_{p,i}| \neq 1$ strictly bounded away from unity, the trace $\mathcal{L}^n$ is

given by the sum over all *periodic points* $i$ of period $n$:

$$\mathrm{tr}\,\mathcal{L}^n = \int dx\,\mathcal{L}^n(x,x) = \sum_{x_i\in \mathrm{Fix} f^n} \frac{e^{\beta\cdot A_i}}{|\det\,(\mathbf{1}-M^n(x_i))|}\,. \tag{16.7}$$

Here Fix $f^n = \{x : f^n(x) = x\}$ is the set of all periodic points of period $n$, and $A_i$ is the observable (15.5) 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 (14.8) by taking the determinant of $\partial_i(x_j - f^n(x)_j)$. If a trajectory retraces itself $r$ times, its fundamental matrix is $M_p^r$, where $M_p$ is the $[d\times d]$ fundamental matrix (4.6) evaluated along a single traversal of the prime cycle $p$. As we saw in (15.5), the integrated observable $A^n$ 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, $A_i = A^n(x_i) = rA_p$, $x_i \in p$.

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 cycle points $\overline{0011} = \overline{1001} = \overline{1100} = \overline{0110}$ belong to the same prime cycle $p = 0011$ of length 4. As both the stability of a cycle and the weight $A_p$ are the same everywhere along the orbit, each prime cycle of length $n_p$ contributes $n_p$ terms to the sum, one for each cycle point. Hence (16.7) can be rewritten as a sum over all prime cycles and their repeats [chapter 10]

$$\mathrm{tr}\,\mathcal{L}^n = \sum_p n_p \sum_{r=1}^{\infty} \frac{e^{r\beta\cdot A_p}}{\left|\det\left(\mathbf{1}-M_p^r\right)\right|}\delta_{n,n_p r}\,, \tag{16.8}$$

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,

$$\sum_{n=1}^{\infty} z^n \mathrm{tr}\,\mathcal{L}^n = \mathrm{tr}\,\frac{z\mathcal{L}}{1-z\mathcal{L}} = \sum_p n_p \sum_{r=1}^{\infty} \frac{z^{n_p r} e^{r\beta\cdot A_p}}{\left|\det\left(\mathbf{1}-M_p^r\right)\right|}\,, \tag{16.9}$$

the constraint $\delta_{n,n_p r}$ is replaced by weight $z^n$. Such discrete time Laplace transform of $\mathrm{tr}\,\mathcal{L}^n$ is usually referred to as a "generating function." Why this transform? We are actually not interested in evaluating the sum (16.8) for any particular fixed period $n$; what we are interested in is the long time $n \to \infty$ behavior. The transform trades in the large time $n$ behavior for the small $z$ behavior. Expressing the trace as in (16.2), in terms of the sum of the eigenvalues of $\mathcal{L}$, we obtain the *trace formula for maps*:

$$\sum_{\alpha=0}^{\infty} \frac{ze^{s_\alpha}}{1-ze^{s_\alpha}} = \sum_p n_p \sum_{r=1}^{\infty} \frac{z^{n_p r}\, e^{r\beta\cdot A_p}}{\left|\det\left(\mathbf{1}-M_p^r\right)\right|}\,. \tag{16.10}$$

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 (13.8).)

fast track:
sect. 16.2, p. 275

**Example 16.1** ***A trace formula for transfer operators:*** *For a piecewise-linear map (15.17), 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 (16.7) reduces to*

$$\mathrm{tr}\,\mathcal{L}^n = \sum_{m=0}^{n}\binom{n}{m}\frac{1}{|1-\Lambda_0^m\Lambda_1^{n-m}|} = \sum_{k=0}^{\infty}\left(\frac{1}{|\Lambda_0|\Lambda_0^k}+\frac{1}{|\Lambda_1|\Lambda_1^k}\right)^n , \tag{16.11}$$

*with eigenvalues*

$$e^{s_k} = \frac{1}{|\Lambda_0|\Lambda_0^k}+\frac{1}{|\Lambda_1|\Lambda_1^k} . \tag{16.12}$$

*As the simplest example of spectrum for such dynamical system, consider the symmetric piecewise-linear 2-branch repeller (15.17) 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 14.7]

*Asymptotically the spectrum (16.12) 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 21.1]

*For $k = 0$ this is in agreement with the explicit transfer matrix (15.19) eigenvalues (15.20). The alert reader should experience anxiety at this point. Is it not true that we have already written down explicitly the transfer operator in (15.19), 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 (15.19) 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 21.1, the full spectrum (16.12) 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 (15.17) follows from (16.9)*

$$\mathrm{tr}\,\frac{z\mathcal{L}}{1-z\mathcal{L}} = \frac{z\left(\frac{1}{|\Lambda_0-1|}+\frac{1}{|\Lambda_1-1|}\right)}{1-z\left(\frac{1}{|\Lambda_0-1|}+\frac{1}{|\Lambda_1-1|}\right)} , \tag{16.13}$$

*verifying the trace formula (16.10).*

## 16.2 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

$$\mathrm{tr}\,\mathcal{L}^t = \mathrm{tr}\,e^{\mathcal{A}t} = \int dx\,\mathcal{L}^t(x,x) = \int dx\,\delta\Big(x - f^t(x)\Big)\,e^{\beta\cdot A^t(x)}\,. \tag{16.14}$$

We are not interested in any particular time $t$, but into the long-time behavior as $t \to \infty$, 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 (14.31). This is a delicate step, since the evolution operator becomes the identity in the $t \to 0^+$ limit. In order to make sense of the trace we regularize the Laplace transform by a lower cutoff $\epsilon$ smaller than the period of any periodic orbit, and write

$$\int_\epsilon^\infty dt\,e^{-st}\,\mathrm{tr}\,\mathcal{L}^t = \mathrm{tr}\,\frac{e^{-(s-\mathcal{A})\epsilon}}{s-\mathcal{A}} = \sum_{\alpha=0}^{\infty}\frac{e^{-(s-s_\alpha)\epsilon}}{s-s_\alpha}\,, \tag{16.15}$$

where $\mathcal{A}$ is the generator of the semigroup of dynamical evolution, see sect. 14.5. Our task is to evaluate $\mathrm{tr}\,\mathcal{L}^t$ from its explicit state space representation.

### 16.2.1 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 fundamental 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 5.2.1]

$$\mathrm{tr}_p\,\mathcal{L}^t = \int_{\mathcal{M}_p} dx_\perp dx_\parallel\,\delta\Big(x_\perp - f_\perp^t(x)\Big)\,\delta\Big(x_\parallel - f^t(x_\parallel)\Big)\ . \tag{16.16}$$

(we set $\beta = 0$ in the $\exp(\beta\cdot A^t)$ weight for the time being). Pick a point on the prime cycle p, and let

$$v(x_\parallel) = \left(\sum_{i=1}^{d} v_i(x)^2\right)^{1/2} \tag{16.17}$$

be the magnitude of the tangential velocity at any point $x = (x_\|, 0, \cdots, 0)$ 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 $0 \le \tau < T_p$, the trajectory $x_\|(\tau) = f^\tau(x_p)$ sweeps out the entire cycle, and for larger times $x_\|$ is a cyclic variable of periodicity $T_p$,

$$x_\|(\tau) = x_\|(\tau + rT_p) \qquad r = 1, 2, \cdots \tag{16.18}$$

We parametrize both the longitudinal coordinate $x_\|(\tau)$ and the velocity $v(\tau) = v(x_\|(\tau))$ by the flight time $\tau$, and rewrite the integral along the periodic orbit as

$$\oint_p dx_\| \, \delta\Big(x_\| - f^t(x_\|)\Big) = \oint_p d\tau \, v(\tau) \, \delta(x_\|(\tau) - x_\|(\tau + t)) \, . \tag{16.19}$$

By the periodicity condition (16.18) the Dirac $\delta$ function picks up contributions for $t = rT_p$, so the Laplace transform can be split as

$$\int_0^\infty dt \, e^{-st} \, \delta(x_\|(\tau) - x_\|(\tau + t)) = \sum_{r=1}^\infty e^{-sT_p r} \, I_r$$
$$I_r = \int_{-\epsilon}^{\epsilon} dt \, e^{-st} \, \delta\Big(x_\|(\tau) - x_\|(\tau + rT_p + t)\Big) \, .$$

Taylor expanding and applying the periodicity condition (16.18), we have $x_\|(\tau + rT_p + t) = x_\|(\tau) + v(\tau)t + \ldots$,

$$I_r = \int_{-\epsilon}^{\epsilon} dt \, e^{-st} \, \delta\Big(x_\|(\tau) - x_\|(\tau + rT_p + t)\Big) = \frac{1}{v(\tau)} \, ,$$

so the remaining integral (16.19) over $\tau$ is simply the cycle period $\oint_p d\tau = T_p$. The contribution of the longitudinal integral to the Laplace transform is thus

$$\int_0^\infty dt \, e^{-st} \oint_p dx_\| \, \delta\Big(x_\| - f^t(x_\|)\Big) = T_p \sum_{r=1}^\infty e^{-sT_p r} \, . \tag{16.20}$$

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.

### 16.2.2 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. 5.3, the transverse stability

eigenvalues do not depend on the choice of a Poincaré section, so ignoring the dependence on $x_{\|}(\tau)$ in evaluating the transverse integral in (16.16) is justified. For the transverse integration variables the fundamental matrix is defined in a reduced Poincaré surface of section $\mathcal{P}$ of fixed $x_{\|}$. Linearization of the periodic flow transverse to the orbit yields

$$\int_{\mathcal{P}} dx_{\perp}\delta\Big(x_{\perp} - f_{\perp}^{rT_p}(x)\Big) = \frac{1}{\left|\det\left(\mathbf{1} - M_p^r\right)\right|} \,, \tag{16.21}$$

where $M_p$ is the $p$-cycle $[d-1\times d-1]$ *transverse* fundamental matrix. As in (16.5) we have to assume hyperbolicity, i.e., that the magnitudes of all transverse eigenvalues are bounded away from unity.

Substitution (16.20), (16.21) in (16.16) leads to an expression for $\mathrm{tr}\,\mathcal{L}^t$ as a sum over all prime cycles $p$ and their repetitions

$$\int_{\epsilon}^{\infty} dt\, e^{-st}\,\mathrm{tr}\,\mathcal{L}^t = \sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r(\beta\cdot A_p - sT_p)}}{\left|\det\left(\mathbf{1} - M_p^r\right)\right|} \,. \tag{16.22}$$

The $\epsilon \to 0$ limit of the two expressions for the resolvent, (16.15) and (16.22), now yields the *classical trace formula for flows*

$$\sum_{\alpha=0}^{\infty} \frac{1}{s - s_\alpha} = \sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r(\beta\cdot A_p - sT_p)}}{\left|\det\left(\mathbf{1} - M_p^r\right)\right|} \,. \tag{16.23}$$

[exercise 16.1]

(If you are worried about the convergence of the resolvent sum, keep the $\varepsilon$ regularization.)

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 $e^{-s} \to z$ throughout, so the trace formula for maps (16.10) is a special case of the trace formula for flows. The relation between the continuous and discrete time cases can be summarized as follows:

$$\begin{aligned} T_p &\leftrightarrow n_p \\ e^{-s} &\leftrightarrow z \\ e^{t\mathcal{A}} &\leftrightarrow \mathcal{L}^n \,. \end{aligned} \tag{16.24}$$

We could now proceed to estimate the location of the leading singularity of $\mathrm{tr}\,(s-\mathcal{A})^{-1}$ by extrapolating finite cycle length truncations of (16.23) 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. 17.2, but first a brief refresher of how all this relates to

the formula for escape rate (1.7) offered in the introduction might help digest the material.

fast track:
sect. 17, p. 283

## 16.3 An asymptotic trace formula

In order to illuminate the manipulations of sect. 16.1 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 (16.10). The Laplace transforms (16.10) or (16.23) are designed to capture the time $\to \infty$ asymptotic behavior of the trace sums. By the hyperbolicity assumption (16.5), for $t = T_p r$ large the cycle weight approaches

$$\left|\det\left(\mathbf{1} - M_p^r\right)\right| \to |\Lambda_p|^r \,, \tag{16.25}$$

where $\Lambda_p$ is the product of the expanding eigenvalues of $M_p$. Denote the corresponding approximation to the $n$th trace (16.7) by

$$\Gamma_n = \sum_i^{(n)} \frac{1}{|\Lambda_i|} \,, \tag{16.26}$$

and denote the approximate trace formula obtained by replacing the cycle weights $\left|\det\left(\mathbf{1} - M_p^r\right)\right|$ by $|\Lambda_p|^r$ in (16.10) by $\Gamma(z)$. Equivalently, think of this as a replacement of the evolution operator (15.23) by a transfer operator (as in example 16.1). 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 (16.8) are labeled by all admissible itineraries composed of sequences of letters $s_i \in \{0, 1\}$:

$$\begin{aligned}
\Gamma(z) &= \sum_{n=1}^{\infty} z^n \Gamma_n = \sum_{n=1}^{\infty} z^n \sum_{x_i \in \mathrm{Fix} f^n} \frac{e^{\beta \cdot A^n(x_i)}}{|\Lambda_i|} \\
&= z\left\{\frac{e^{\beta\cdot A_0}}{|\Lambda_0|} + \frac{e^{\beta\cdot A_1}}{|\Lambda_1|}\right\} + z^2\left\{\frac{e^{2\beta\cdot A_0}}{|\Lambda_0|^2} + \frac{e^{\beta\cdot A_{01}}}{|\Lambda_{01}|} + \frac{e^{\beta\cdot A_{10}}}{|\Lambda_{10}|} + \frac{e^{2\beta\cdot A_1}}{|\Lambda_1|^2}\right\} \\
&\quad + z^3\left\{\frac{e^{3\beta\cdot A_0}}{|\Lambda_0|^3} + \frac{e^{\beta\cdot A_{001}}}{|\Lambda_{001}|} + \frac{e^{\beta\cdot A_{010}}}{|\Lambda_{010}|} + \frac{e^{\beta\cdot A_{100}}}{|\Lambda_{100}|} + \ldots\right\}
\end{aligned} \tag{16.27}$$

Both the cycle averages $A_i$ and the stabilities $\Lambda_i$ are the same for all points $x_i \in p$ in a cycle $p$. Summing over repeats of all prime cycles we obtain

$$\Gamma(z) = \sum_p \frac{n_p t_p}{1 - t_p} \,, \qquad t_p = z^{n_p} e^{\beta\cdot A_p}/|\Lambda_p| \,. \tag{16.28}$$

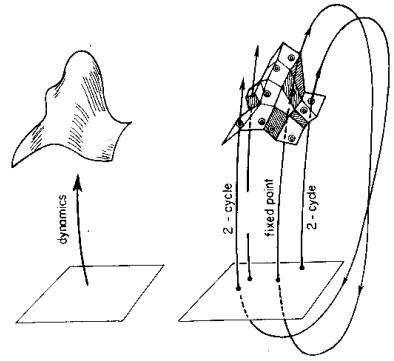



**Figure 16.1:** Approximation to (a) a smooth dynamics by (b) the skeleton of periodic points, together with their linearized neighborhoods. Indicated are segments of two 1-cycles and a 2-cycle that alternates between the neighborhoods of the two 1-cycles, shadowing first one of the two 1-cycles, and then the other.

This is precisely our initial heuristic estimate (1.8). Note that we could not perform such sum over $r$ in the exact trace formula (16.10) as $\left|\det\left(\mathbf{1}-M_p^r\right)\right| \neq \left|\det\left(\mathbf{1}-M_p\right)\right|^r$; 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 (17.9). [section 17.2]

If the weights $e^{\beta A^n(x)}$ are multiplicative along the flow, and the flow is hyperbolic, for given $\beta$ the magnitude of each $|e^{\beta A^n(x_i)}/\Lambda_i|$ 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 (16.7) tends to the leading $\mathcal{L}^n$ eigenvalue $e^{ns_0}$. Summing this asymptotic estimate level by level

$$\Gamma(z) \approx \sum_{n=1}^{\infty} (ze^{s_0})^n = \frac{ze^{s_0}}{1-ze^{s_0}} \tag{16.29}$$

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 (16.28) 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. 17.3 to motivate the introduction of dynamical zeta functions (17.14), and in sect. 17.5 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 16.1, with a smooth flow approximated by its *periodic orbit skeleton*, each region $\mathcal{M}_i$ centered on a periodic point $x_i$ of the topological length $n$, 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 *classical trace formula*

$$\sum_{\alpha=0}^{\infty} \frac{1}{s-s_\alpha} = \sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r(\beta\cdot A_p - sT_p)}}{\left|\det\left(\mathbf{1}-M_p^r\right)\right|} .$$

Now that we have a trace formula, we might ask for what is it good? As it stands, it is little more than 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 21). 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 chapter11. The more onerous enterprize of actually computing the cycles we first approach traditionally, as a numerical task in chapter 12, and then more boldly as a part and parcel of variational foundations of classical and quantum dynamics in chapter27.

## Commentary

**Remark 16.1** Who's dunne it? Continuous time flow traces weighted by cycle periods were introduced by Bowen [1] who treated them as Poincaré section suspensions weighted by the "time ceiling" function (3.5). They were used by Parry and Pollicott [2].

**Remark 16.2** 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 (16.14) 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 (16.7) are called *flat* traces (see the review [4] and chapter 21). Other names for traces appear as well: for instance, in the context of 1-$d$ mappings, *sharp* traces refer to generalizations of (16.7) where contributions of periodic points are weighted by the Lefschetz sign $\pm 1$, 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 [11]. Such traces are connected to the theory of kneading invariants (see ref. [4] and references therein). Traces weighted by $\pm 1$ 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 $\delta$, refs. [5, 6, 6].

## Exercises

16.1. $t \to 0_+$ **regularization of eigenvalue sums**$^{**}$**.** In taking the Laplace transform (16.23) we have ignored the $t \to 0_+$ 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 (16.23) and assign to such volume term some interesting role in the theory of classical resonance spectra. E-mail the solution to the authors.

16.2. **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 $\mathbb{R}^{n_1} \mapsto \mathbb{R}^{n_2}$? ($n_i \in \mathbb{N}$.)

## References


[16.1] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Springer Lecture Notes in Math. **470** (1975).

[16.2] W. Parry and M. Pollicott, *Zeta Functions and the periodic Structure of Hyperbolic Dynamics*, *Astérisque* **187–188** (Société Mathématique de France, Paris 1990).

[16.3] P. Cvitanović and B. Eckhardt, *J. Phys.* **A 24**, L237 (1991).

[16.4] V. Baladi and D. Ruelle, *Ergodic Theory Dynamical Systems* **14**, 621 (1994).

[16.5] R. Artuso, E. Aurell and P. Cvitanović, *Nonlinearity* **3**, 325 (1990); 361 (1990).

[16.6] M. Pollicott, *J. Stat. Phys.* **62**, 257 (1991).

# Chapter 17

# 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 (16.10), (16.23) and (16.28) 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. The idea is illustrated by figure 1.13: Determinants tend to have larger analyticity domains because if $\mathrm{tr}\,\mathcal{L}/(1 - z\mathcal{L}) = -\frac{d}{dz}\ln\det(1 - z\mathcal{L})$ 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 21]

## 17.1 Spectral determinants for maps

The eigenvalues $z_k$ of a linear operator are given by the zeros of the determinant

$$\det(1 - z\mathcal{L}) = \prod_k (1 - z/z_k)\,. \tag{17.1}$$

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]

$$\ln\det(1-M) = \mathrm{tr}\,\ln(1-M) = -\sum_{n=1}^{\infty}\frac{1}{n}\mathrm{tr}\,M^n\,, \tag{17.2}$$

to relate the spectral determinant of an evolution operator for a map to its traces (16.8), and hence to periodic orbits:

$$\begin{aligned}\det(1-z\mathcal{L}) &= \exp\left(-\sum_{n}^{\infty}\frac{z^n}{n}\mathrm{tr}\,\mathcal{L}^n\right)\\ &= \exp\left(-\sum_{p}\sum_{r=1}^{\infty}\frac{1}{r}\frac{z^{n_p r}e^{r\beta\cdot A_p}}{\left|\det\left(\mathbf{1}-M_p^r\right)\right|}\right)\,.\end{aligned} \tag{17.3}$$

Going the other way, the trace formula (16.10) can be recovered from the spectral determinant by taking a derivative

$$\mathrm{tr}\,\frac{z\mathcal{L}}{1-z\mathcal{L}} = -z\frac{d}{dz}\ln\det(1-z\mathcal{L})\,. \tag{17.4}$$

fast track:
sect. 17.2, p. 285

**Example 17.1** ***Spectral determinants of transfer operators:***

*For a piecewise-linear map (15.17) with a finite Markov partition, an explicit formula for the spectral determinant follows by substituting the trace formula (16.11) into (17.3):*

$$\det(1-z\mathcal{L}) = \prod_{k=0}^{\infty}\left(1-\frac{t_0}{\Lambda_0^k}-\frac{t_1}{\Lambda_1^k}\right)\,, \tag{17.5}$$

*where $t_s = z/|\Lambda_s|$. The eigenvalues are necessarily the same as in (16.12), which we already determined from the trace formula (16.10).*

*The exponential spacing of eigenvalues guarantees that the spectral determinant (17.5) 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.*

## 17.2 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 $\gamma$ [···] then define formally (another zeta function!) $Z(s)$ to be the infinite product
>
> $$Z(s) = \prod_{\gamma\in\Gamma}\prod_{k=0}^{\infty}\left(1-[\exp l(\gamma)]^{-s-k}\right).$$
>
> —Stephen Smale, *Differentiable Dynamical Systems*

We write the formula for the spectral determinant for flows by analogy to (17.3)

$$\det(s-\mathcal{A}) = \exp\left(-\sum_p\sum_{r=1}^{\infty}\frac{1}{r}\frac{e^{r(\beta\cdot A_p - sT_p)}}{\left|\det\left(\mathbf{1}-M_p^r\right)\right|}\right), \tag{17.6}$$

and then check that the trace formula (16.23) is the logarithmic derivative of the spectral determinant

$$\mathrm{tr}\,\frac{1}{s-\mathcal{A}} = \frac{d}{ds}\ln\det(s-\mathcal{A}). \tag{17.7}$$

With $z$ set to $z = e^{-s}$ as in (16.24), the spectral determinant (17.6) has the same form for both maps and flows. We refer to (17.6) as *spectral determinant*, as the spectrum of the operator $\mathcal{A}$ is given by the zeros of

$$\det(s-\mathcal{A}) = 0. \tag{17.8}$$

We now note that the $r$ sum in (17.6) 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 fundamental matrix, with eigenvalues $\Lambda_{p,1}, \Lambda_{p,2}, \ldots, \Lambda_{p,d}$. Expanding the expanding eigenvalue factors $1/(1-1/\Lambda_{p,e})$ and the contracting eigenvalue factors $1/(1-\Lambda_{p,c})$ in (16.4) as geometric series, substituting back into (17.6), and resumming the logarithms, we find that the spectral determinant is formally given by the infinite product

$$\det(s-\mathcal{A}) = \prod_{k_1=0}^{\infty}\cdots\prod_{l_c=0}^{\infty}\frac{1}{\zeta_{k_1\cdots l_c}}$$

$$1/\zeta_{k_1\cdots l_c} = \prod_p\left(1-t_p\frac{\Lambda_{p,e+1}^{l_1}\Lambda_{p,e+2}^{l_2}\cdots\Lambda_{p,d}^{l_c}}{\Lambda_{p,1}^{k_1}\Lambda_{p,2}^{k_2}\cdots\Lambda_{p,e}^{k_e}}\right) \tag{17.9}$$

$$t_p = t_p(z,s,\beta) = \frac{1}{|\Lambda_p|}e^{\beta\cdot A_p - sT_p}z^{n_p}. \tag{17.10}$$

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 (16.10) for maps, whereas the factor $e^{-sT_p}$ is specific to the continuous time trace formuls (16.23); according to (16.24) 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 lengths. These factors will assist us in expanding zeta functions and determinants, eventually we shall set $z = 1$. [chapter 18] 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 $A^t(x)$ term in the $p$ cycle multiplicative weight $e^{\beta \cdot A_p}$. By its definition (15.1), the weight for maps is a product along the cycle points

$$e^{A_p} = \prod_{j=0}^{n_p-1} e^{a(f^j(x_p))} ,$$

and the weight for flows is an exponential of the integral (15.5) 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 fundamental matrix of sect. 4.1.

**Example 17.2** ***Expanding 1-d map:*** *For expanding 1-d mappings the spectral determinant (17.9) takes the form*

$$\det(1 - z\mathcal{L}) = \prod_p \prod_{k=0}^{\infty} \left(1 - t_p/\Lambda_p^k\right) , \qquad t_p = \frac{e^{\beta A_p}}{|\Lambda_p|} z^{n_p} . \tag{17.11}$$

**Example 17.3** ***Two-degree of freedom Hamiltonian flows:*** *For a 2-degree of freedom Hamiltonian flows the energy conservation eliminates on phase space variable, and restriction to a Poincaré section eliminates the marginal longitudinal eigenvalue $\Lambda = 1$, so a periodic orbit of 2-degree of freedom hyperbolic Hamiltonian flow has one expanding transverse eigenvalue $\Lambda$, $|\Lambda| > 1$, and one contracting transverse eigenvalue $1/\Lambda$. The weight in (16.4) is expanded as follows:*

$$\frac{1}{\left|\det\left(\mathbf{1} - M_p^r\right)\right|} = \frac{1}{|\Lambda|^r (1 - 1/\Lambda_p^r)^2} = \frac{1}{|\Lambda|^r} \sum_{k=0}^{\infty} \frac{k+1}{\Lambda_p^{kr}} . \tag{17.12}$$

*The spectral determinant exponent can be resummed,*

$$-\sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{(\beta A_p - sT_p)r}}{\left|\det\left(\mathbf{1} - M_p^r\right)\right|} = \sum_{k=0}^{\infty} (k+1) \log\left(1 - \frac{e^{\beta A_p - sT_p}}{|\Lambda_p|\Lambda_p^k}\right) ,$$

*and the spectral determinant for a 2-dimensional hyperbolic Hamiltonian flow rewritten as an infinite product over prime cycles*

$$\det(s-\mathcal{A}) = \prod_p \prod_{k=0}^{\infty} \left(1 - t_p/\Lambda_p^k\right)^{k+1} \quad . \tag{17.13}$$

[exercise 21.4]

Now we are finally poised to deal with the problem posed at the beginning of chapter 16; how do we actually evaluate the averages introduced in sect.15.1? The eigenvalues of the dynamical averaging evolution operator are given by the values of $s$ for which the spectral determinant (17.6) of the evolution operator (15.23) vanishes. If we can compute the leading eigenvalue $s_0(\beta)$ and its derivatives, we are done. Unfortunately, the infinite product formula (17.9) 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 18, but here it is natural to introduce still another variant of a determinant, the dynamical zeta function.

## 17.3 Dynamical zeta functions

It follows from sect. 16.1.1 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 $rT_p = t \to \infty$ limit, where $\Lambda_p = \prod_e \Lambda_{p,e}$ is the product of the expanding eigenvalues of the fundamental matrix $M_p$. With this replacement the spectral determinant (17.6) is replaced by the *dynamical zeta function*

$$1/\zeta = \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{1}{r} t_p^r\right) \tag{17.14}$$

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:

$$1/\zeta = \prod_p \left(1 - t_p\right) . \tag{17.15}$$

In order to simplify the notation, we usually omit the explicit dependence of $1/\zeta$, $t_p$ on $z, s, \beta$ whenever the dependence is clear from the context.

The approximate trace formula (16.28) plays the same role *vis-à-vis* the dynamical zeta function (17.7)

$$\Gamma(s) = \frac{d}{ds} \ln \zeta^{-1} = \sum_p \frac{T_p t_p}{1 - t_p} , \tag{17.16}$$

as the exact trace formula (16.23) plays *vis-à-vis* the spectral determinant (17.6). The heuristically derived dynamical zeta function of sect. 1.5.2 now re-emerges as the $1/\zeta_{0\cdots0}(z)$ part of the *exact* spectral determinant; other factors in the infinite product (17.9) affect the non-leading eigenvalues of $\mathcal{L}$.

In summary, the dynamical zeta function (17.15) associated with the flow $f^t(x)$ is defined as the product over all prime cycles $p$. The quantities, $T_p$, $n_p$ and $\Lambda_p$, denote the period, topological length and product of the expanding stability eigenvalues of prime cycle $p$, $A_p$ is the integrated observable $a(x)$ evaluated on a single traversal of cycle $p$ (see (15.5)), $s$ 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

$$1/\zeta(s) = 0 \tag{17.17}$$

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 classical 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 18, p. 299

### 17.3.1 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 23): $\Gamma_n$, the trace sum (16.26), can be expressed in terms of the dynamical zeta function (17.15)

$$1/\zeta(z) = \prod_p \left(1 - \frac{z^{n_p}}{|\Lambda_p|}\right) \quad . \tag{17.18}$$

as a contour integral

$$\Gamma_n = \frac{1}{2\pi i} \oint_{\gamma_r^-} z^{-n} \left(\frac{d}{dz} \log \zeta^{-1}(z)\right) dz \quad , \tag{17.19}$$

[exercise 17.7]

where a small contour $\gamma_r^-$ 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

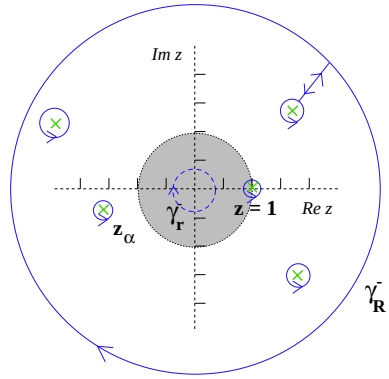



**Figure 17.1:** The survival probability $\Gamma_n$ can be split into contributions from poles (x) and zeros (o) between the small and the large circle and a contribution from the large circle.

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 (16.26) 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 17.1. Residue calculus turns this into a sum over the zeros $z_\alpha$ and poles $z_\beta$ of the dynamical zeta function, that is

[chapter 18]

$$\Gamma_n = \sum_{|z_\alpha|<R}^{\text{zeros}} \frac{1}{z_\alpha^n} - \sum_{|z_\beta|<R}^{\text{poles}} \frac{1}{z_\beta^n} + \frac{1}{2\pi i} \oint_{\gamma_R^-} dz\, z^{-n} \frac{d}{dz} \log \zeta^{-1}, \tag{17.20}$$

where the last term gives a contribution from a large circle $\gamma_R^-$. It would be a miracle if you still remembered this, but in sect. 1.4.3 we interpreted $\Gamma_n$ as fraction of survivors after $n$ bounces, and defined the escape rate $\gamma$ as the rate of the find exponential decay of $\Gamma_n$. We now see that this exponential decay is dominated by the leading zero or pole of $\zeta^{-1}(z)$.

### 17.3.2 Dynamical zeta functions for transfer operators

Ruelle's original dynamical zeta function was a generalization of the topological zeta function (13.21) to a function that assigns different weights to different cycles:

[chapter 13]

$$\zeta(z) = \exp \sum_{n=1}^{\infty} \frac{z^n}{n} \left( \sum_{x_i \in \mathrm{Fix}\, f^n} \mathrm{tr} \prod_{j=0}^{n-1} g(f^j(x_i)) \right).$$

[exercise 16.2]

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 $x_i$.

By the chain rule (4.50) the stability of any $n$-cycle of a 1-$d$ map is given by $\Lambda_p = \prod_{j=1}^{n} f'(x_i)$, so the 1-$d$ 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 (14.9) - 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 14.1, the dynamical zeta function is given by a finite polynomial, a straightforward generalization of the topological transition matrix determinant (10.2). As explained in sect. 13.3, 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 17.4** ***A piecewise linear repeller:*** *Due to piecewise linearity, the stability of any $n$-cycle of the piecewise linear repeller (15.17) factorizes as $\Lambda_{s_1 s_2 \ldots s_n} = \Lambda_0^m \Lambda_1^{n-m}$, 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 (16.28) 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 (17.14) evaluated by resumming the traces,*

[exercise 17.3]

$$1/\zeta(z) = 1 - z/|\Lambda_0| - z/|\Lambda_1| \,, \tag{17.21}$$

*is indeed the determinant* $\det(1-zT)$ *of the transfer operator (15.19), which is almost as simple as the topological zeta function (13.25).*

[section 10.5]

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 chaologians 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 shall learn in chapter 18 how to exploit this redundancy systematically.

## 17.4 False zeros

Compare (17.21) with the Euler product (17.15). 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

(17.15) vanishes whenever one of its factors vanishes. If that were true, each factor $(1 - z^{n_p}/|\Lambda_p|)$ would yield

$$0 = 1 - e^{n_p(\gamma - \lambda_p)}, \tag{17.22}$$

so the escape rate $\gamma$ would equal the Floquet exponent of a repulsive 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 (17.21)

$$0 = 1 - e^{\gamma - \lambda + h}, \qquad h = \ln 2. \tag{17.23}$$

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 into this book. 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 (17.15) to the true zero $z = e^{\lambda - h}$.

## 17.5 Spectral determinants *vs.* dynamical zeta functions

In sect. 17.3 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

$$1 = \frac{1}{\det(1 - M)} \sum_{k=0}^{d} (-1)^k \mathrm{tr}\left(\wedge^k M\right), \tag{17.24}$$

inserted into the exponential representation (17.14) of the dynamical zeta function, relates the dynamical zeta function to *weighted* spectral determinants.

**Example 17.5** ***Dynamical zeta function in terms of determinants, 1-d maps:*** *For 1-d maps the identity*

$$1 = \frac{1}{(1 - 1/\Lambda)} - \frac{1}{\Lambda}\frac{1}{(1 - 1/\Lambda)}$$

*substituted into (17.14) yields an expression for the dynamical zeta function for 1-d maps as a ratio of two spectral determinants*

$$1/\zeta = \frac{\det(1 - z\mathcal{L})}{\det(1 - z\mathcal{L}_{(1)})} \tag{17.25}$$

*where the cycle weight in $\mathcal{L}_{(1)}$ is given by replacement $t_p \to t_p/\Lambda_p$. As we shall see in chapter 21, 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 $\det(1-z\mathcal{L}_{(1)})$. Hence even though we have derived the dynamical zeta function in sect. 17.3 as an "approximation" to the spectral determinant, the two contain the same spectral information.*

**Example 17.6** ***Dynamical zeta function in terms of determinants, 2-$d$ 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*

$$1/\zeta = \frac{\det(1-z\mathcal{L})\det(1-z\mathcal{L}_{(2)})}{\det(1-z\mathcal{L}_{(1)})}\,. \tag{17.26}$$

*This establishes that for nice 2-$d$ hyperbolic flows the dynamical zeta function is meromorphic.*

**Example 17.7** ***Dynamical zeta functions for 2-$d$ Hamiltonian flows:*** *The relation (17.26) 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 (17.14) of $1/\zeta_k$, and obtain*

$$1/\zeta_k = \frac{\det(1-z\mathcal{L}_{(k)})\det(1-z\mathcal{L}_{(k+2)})}{\det(1-z\mathcal{L}_{(k+1)})^2}\,. \tag{17.27}$$

*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 (17.13) 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.*

## 17.6 All too many eigenvalues?

What does the 2-dimensional hyperbolic Hamiltonian flow spectral determinant (17.13) tell us? Consider one of the simplest conceivable hyperbolic flows: the game of pinball of figure **??** consisting of two disks of equal size in a plane. There

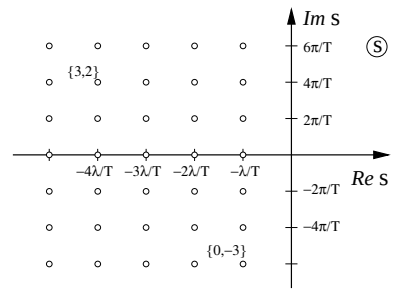



**Figure 17.2:** The classical resonances $\alpha = \{k, n\}$ (17.28) for a 2-disk game of pinball.

is only one periodic orbit, with the period $T$ and expanding eigenvalue $\Lambda$ given by elementary considerations (see exercise 9.3), and the resonances $\det(s_\alpha - \mathcal{A}) = 0$, $\alpha = \{k, n\}$ plotted in figure 17.2:

$$s_\alpha = -(k+1)\lambda + n\frac{2\pi i}{T}, \quad n \in \mathbb{Z}, \, k \in \mathbb{Z}_+, \quad \text{multiplicity } k+1, \tag{17.28}$$

can be read off the spectral determinant (17.13) for a single unstable cycle:

$$\det(s - \mathcal{A}) = \prod_{k=0}^{\infty} \left(1 - e^{-sT}/|\Lambda|\Lambda^k\right)^{k+1} . \tag{17.29}$$

In the above $\lambda = \ln|\Lambda|/T$ is the cycle Lyapunov exponent. For an open system, the real part of the eigenvalue $s_\alpha$ gives the decay rate of $\alpha$th eigenstate, and the imaginary part gives the "node number" of the eigenstate. The negative real part of $s_\alpha$ 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 $\mathrm{Re}\, s_\alpha$ 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 (16.23), the trace formula for flows:

$$\begin{aligned}
\sum_{\alpha=0}^{\infty} e^{s_\alpha t} &= \sum_{k=0}^{\infty} \sum_{n=-\infty}^{\infty} (k+1) e^{(k+1)\lambda t + i2\pi n t/T} \\
&= \sum_{k=0}^{\infty} (k+1) \left(\frac{1}{|\Lambda|\Lambda^k}\right)^{t/T} \sum_{n=-\infty}^{\infty} e^{i2\pi n/T} \\
&= \sum_{k=0}^{\infty} \frac{k+1}{|\Lambda|^r \Lambda^{kr}} \sum_{r=-\infty}^{\infty} \delta(r - t/T) \\
&= T \sum_{r=-\infty}^{\infty} \frac{\delta(t - rT)}{|\Lambda|(1 - 1/\Lambda^r)^2} .
\end{aligned} \tag{17.30}$$

Hence, the two sides of the trace formula (16.23) are verified. The formula is fine for $t > 0$; for $t \to 0_+$, 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 $\mathrm{Im}\, s$ 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 (16.8). 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 18.1.

## 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 $\log\det = \mathrm{tr}\log$. 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*

$$\det(s-\mathcal{A}) = \exp\left(-\sum_p\sum_{r=1}^{\infty}\frac{1}{r}\frac{e^{(\beta\cdot A_p - sT_p)r}}{\left|\det\left(\mathbf{1}-M_p^r\right)\right|}\right),$$

or the leading zeros of the *dynamical zeta function*

$$1/\zeta = \prod_p\left(1-t_p\right), \qquad t_p = \frac{1}{|\Lambda_p|}e^{\beta\cdot A_p - sT_p}.$$

The spectral determinant is the tool of choice in actual calculations, as it has superior convergence properties (this will be discussed in chapter 21 and is illustrated, for example, by table **??**). 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 $s_\alpha$, and are analytic in $s$ in an open neighborhood of $s_\alpha$.

The critical step in the derivation of the periodic orbit formulas for spectral determinants and dynamical zeta functions is the hyperbolicity assumption (16.5) that no cycle stability eigenvalue is marginal, $|\Lambda_{p,i}| \neq 1$. By dropping the prefactors in (1.4), 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 18) 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 17.1** 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 [24]: expanding holomorphic case, see Ruelle [9]; piecewise monotone maps of the interval, see Hofbauer and Keller [14] and Baladi and Keller [17].

**Remark 17.2** Smale's wild idea. Smale's wild idea quoted on page 285 was technically wrong because 1) the Selberg zeta function yields the spectrum of a quantum mechanical Laplacian rather than the classical 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 $e^{-sT_p}/\Lambda_p^k$ in (17.9) simplify to $s^{-(s+k)T_p}$, as intuited in Smale's quote on page 285 (where $l(\gamma)$ is the cycle period denoted here by $T_p$). Nevertheless, Smale's intuition was remarkably on the target.

**Remark 17.3** 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. [3], a very readable introduction to the Selberg Zeta function.

**Remark 17.4** Zeta functions, antecedents. For a function to be deserving of the appellation "zeta function," one expects it to have an Euler product representation (17.15), and perhaps also satisfy a functional equation. Various kinds of zeta functions are reviewed in refs. [7, 8, 9]. Historical antecedents of the dynamical zeta function are the fixed-point counting functions introduced by Weil [10], Lefschetz [11] and Artin and Mazur [12], and the determinants of transfer operators of statistical mechanics [26].

In his review article Smale [23] already intuited, by analogy to the Selberg Zeta function, that the spectral determinant is the right generalization for continuous time flows. 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 (17.3) between spectral determinants and dynamical zeta functions, but since he was motivated by the Artin-Mazur zeta function (13.21) 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" [18, 23].

The nomenclature has not settled down yet; what we call evolution operators here is elsewhere called transfer operators [28], Perron-Frobenius operators [5] and/or Ruelle-Araki operators.

Here we refer to kernels such as (15.23) as evolution operators. We follow Ruelle in usage of the term "dynamical zeta function," but elsewhere in the literature the function (17.15) is often called the Ruelle zeta function. Ruelle [29] 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], functional determinants, Fredholm determinants, or even - to maximize confusion - dynamical zeta functions [13]. 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."

## Exercises

17.1. **Escape rate for a 1-$d$ repeller, numerically.** Consider the quadratic map

$$f(x) = Ax(1-x) \tag{17.31}$$

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 (20.8), 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 18.2.

17.2. **Spectrum of the "golden mean" pruned map.** (medium - Exercise 13.6 continued)

(a) Determine an expression for $\mathrm{tr}\,\mathcal{L}^n$, the trace of powers of the Perron-Frobenius operator (14.10) for the tent map of exercise 13.6.

(b) Show that the spectral determinant for the Perron-Frobenius operator is

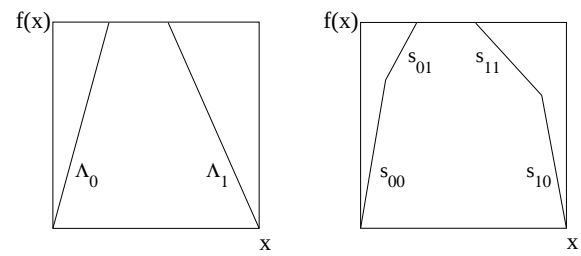
$$\det(1 - z\mathcal{L}) = \prod_{k \text{ even}} \left(1 - \frac{z}{\Lambda^{k+1}} - \frac{z^2}{\Lambda^{2k+2}}\right) \prod_{k \text{ odd}} \left(1 + \frac{z}{\Lambda^{k+1}} + \frac{z^2}{\Lambda^{2k+2}}\right) \tag{17.32}$$

17.3. **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 (15.17) with the left branch slope $\Lambda_0$, the right branch slope $\Lambda_1$.




(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?

17.4. **Dynamical zeta functions from Markov graphs.** Extend sect. 13.3 to evaluation of dynamical zeta functions for piecewise linear maps with finite Markov graphs. This generalizes the results of exercise 17.3.

17.5. **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

$$F(z) = \prod_p (1 - \frac{z^{n_p}}{\Lambda^{n_p}})$$

(e) Are the roots of the factors in the above product the zeros of $F(z)$?

(Per Rosenqvist)

17.6. **Dynamical zeta functions as ratios of spectral determinants.** (medium) Show that the zeta function

$$1/\zeta(z) = \exp\left(-\sum_p \sum_{r=1} \frac{1}{r} \frac{z^{n_p}}{|\Lambda_p|^r}\right)$$

can be written as the ratio $1/\zeta(z) = \det(1 - z\mathcal{L}_{(0)})/\det(1 - z\mathcal{L}_{(1)})$, where $\det(1 - z\mathcal{L}_{(s)}) = \prod_p \prod_{k=0}^{\infty}(1 - z^{n_p}/|\Lambda_p|\Lambda_p^{k+s})$.

17.7. **Contour integral for survival probability.** Perform explicitly the contour integral appearing in (17.19).

17.8. **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 Ulam tent map (14.45)

$$1/\zeta(z) = \prod_{p\in P}\left(1 - \frac{z^{n_p}}{|\Lambda_p|}\right).$$

What are its roots? Do they agree with those computed in exercise 14.7?

17.9. **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 $t_p \to t_p/\Lambda_p^k$ in the spectral determinant.

17.10. **Riemann $\zeta$ function.** The Riemann $\zeta$ function is defined as the sum

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \qquad s \in \mathbb{C}.$$

(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 17.15 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 $\zeta$ function? If not, why not?

17.11. **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:

$$\Lambda_p = \Lambda_0^{n_{p,0}}\Lambda_1^{n_{p,1}}, \qquad n_{p,0} = \#0 \text{ in } p, \; n_{p,1} = \#1 \text{ in } p,$$

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 (15.19), 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 length with $n_p \le N$, 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 length $N$?

## References

[17.1] D. Ruelle, *Statistical Mechanics, Thermodynamic Formalism* (Addison-Wesley, Reading MA, 1978)

[17.2] M. Pollicott, "Meromorphic extensions of generalised zeta functions," *Invent. Math.* **85**, 147 (1986).

[17.3] H.P. McKean, *Comm. Pure and Appl. Math.* **25** , 225 (1972); **27**, 134 (1974).

[17.4] W. Parry and M. Pollicott, *Ann. Math.* **118**, 573 (1983).

[17.5] Y. Oono and Y. Takahashi, *Progr. Theor. Phys* **63**, 1804 (1980); S.-J. Chang and J. Wright, *Phys. Rev. A* **23**, 1419 (1981); Y. Takahashi and Y. Oono, *Progr. Theor. Phys* **71**, 851 (1984).

[17.6] P. Cvitanović, P.E. Rosenqvist, H.H. Rugh, and G. Vattay, "A Fredholm determinant for semi-classical quantization," *CHAOS* **3**, 619 (1993).

[17.7] A. Voros, in: *Zeta Functions in Geometry* (Proceedings, Tokyo 1990), eds. N. Kurokawa and T. Sunada, Advanced Studies in Pure Mathematics **21**, Math. Soc. Japan, Kinokuniya, Tokyo (1992), p.327-358.

[17.8] Kiyosi Itô, ed., *Encyclopedic Dictionary of Mathematics*, (MIT Press, Cambridge, 1987).

[17.9] N.E. Hurt, "Zeta functions and periodic orbit theory: A review," *Results in Mathematics* **23**, 55 (Birkhäuser, Basel 1993).

[17.10] A. Weil, "Numbers of solutions of equations in finite fields," *Bull. Am. Math. Soc.* **55**, 497 (1949).

[17.11] D. Fried, "Lefschetz formula for flows," *The Lefschetz centennial conference*, *Contemp. Math.* **58**, 19 (1987).

[17.12] E. Artin and B. Mazur, *Annals. Math.* **81**, 82 (1965)

[17.13] M. Sieber and F. Steiner, *Phys. Lett.* **A 148**, 415 (1990).

[17.14] F. Hofbauer and G. Keller, "Ergodic properties of invariant measures for piecewise monotonic transformations," *Math. Z.* **180**, 119 (1982).

[17.15] G. Keller, "On the rate of convergence to equilibrium in one-dimensional systems," *Comm. Math. Phys.* **96**, 181 (1984).

[17.16] F. Hofbauer and G. Keller, "Zeta-functions and transfer-operators for piecewise linear transformations," *J. reine angew. Math.* **352**, 100 (1984).

[17.17] V. Baladi and G. Keller, "Zeta functions and transfer operators for piecewise monotone transformations," *Comm. Math. Phys.* **127**, 459 (1990).

# Chapter 18

# Cycle expansions

Recycle... It's the Law!
—Poster, New York City Department of Sanitation

THE EULER PRODUCT representations of spectral determinants (17.9) and dynamical zeta functions (17.15) are really only a shorthand notation - the zeros of the individual factors are *not* the zeros of the zeta function, and convergence of such objects is far from obvious. Now we shall give meaning to the dynamical zeta functions and spectral determinants by expanding them as cycle expansions, series representations ordered by increasing topological cycle length, with products in (17.9), (17.15) expanded as sums over *pseudocycles*, products of $t_p$'s. The zeros of correctly truncated cycle expansions yield the desired eigenvalues, 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).

## 18.1 Pseudocycles and shadowing

How are periodic orbit formulas such as (17.15) evaluated? We start by computing the lengths and stability eigenvalues of the shortest cycles. This always requires numerical work, such as the Newton method searches for periodic solutions; we shall assume that the numerics is under control, and that *all* short cycles up to a given (topological) length have been found. Examples of the data required for application of periodic orbit formulas are the lists of cycles given in table **??** and exercise 12.11. 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, improves it, but painfully slowly).

Expand the dynamical zeta function (17.15) as a formal power series,

$$1/\zeta = \prod_p (1 - t_p) = 1 - \sum_{\{p_1 p_2 \cdots p_k\}}{}' (-1)^{k+1} t_{p_1} t_{p_2} \cdots t_{p_k} \tag{18.1}$$

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 $t_p$. The formal power series (18.1) is now compactly written as

$$1/\zeta = 1 - \sum_\pi{}' \, t_\pi \, . \tag{18.2}$$

For $k > 1$, $t_\pi$ are weights of *pseudocycles*; they are sequences of shorter cycles that shadow a cycle with the symbol sequence $p_1 p_2 \dots p_k$ along segments $p_1$, $p_2, \dots, p_k$. $\sum'$ denotes the restricted sum, for which any given prime cycle $p$ contributes at most once to a given pseudocycle weight $t_\pi$.

The pseudocycle weight, i.e., the product of weights (17.10) of prime cycles comprising the pseudocycle,

$$t_\pi = (-1)^{k+1} \frac{1}{|\Lambda_\pi|} e^{\beta A_\pi - s T_\pi} z^{n_\pi} \, , \tag{18.3}$$

depends on the pseudocycle topological length $n_\pi$, integrated observable $A_\pi$, period $T_\pi$, and stability $\Lambda_\pi$

$$\begin{aligned} n_\pi &= n_{p_1} + \dots + n_{p_k} \, , \qquad T_\pi = T_{p_1} + \dots + T_{p_k} \\ A_\pi &= A_{p_1} + \dots + A_{p_k} \, , \qquad \Lambda_\pi = \Lambda_{p_1} \Lambda_{p_2} \cdots \Lambda_{p_k} \, . \end{aligned} \tag{18.4}$$

Throughout this text, the terms "periodic orbit" and "cycle" are used interchangeably; while "periodic orbit" is more precise, "cycle" (which has many other uses in mathematics) is easier on the ear than "pseudo-periodic-orbit." While in Soviet times acronyms were a rage (and in France they remain so), we shy away from acronyms such as UPOs (Unstable Periodic Orbits).

### 18.1.1 Curvature expansions

The simplest example is the pseudocycle sum for a system described by a complete binary symbolic dynamics. In this case the Euler product (17.15) is given by

$$\begin{aligned} 1/\zeta \;=\; & (1-t_0)(1-t_1)(1-t_{01})(1-t_{001})(1-t_{011}) \\ & (1-t_{0001})(1-t_{0011})(1-t_{0111})(1-t_{00001})(1-t_{00011}) \\ & (1-t_{00101})(1-t_{00111})(1-t_{01011})(1-t_{01111}) \dots \end{aligned} \tag{18.5}$$

(see table **??**), and the first few terms of the expansion (18.2) ordered by increasing total pseudocycle length are:

$$\begin{aligned} 1/\zeta \;=\; & 1 - t_0 - t_1 - t_{01} - t_{001} - t_{011} - t_{0001} - t_{0011} - t_{0111} - \dots \\ & + t_0 t_1 + t_0 t_{01} + t_{01} t_1 + t_0 t_{001} + t_0 t_{011} + t_{001} t_1 + t_{011} t_1 \\ & - t_0 t_{01} t_1 - \dots \end{aligned} \tag{18.6}$$

We refer to such series representation of a dynamical zeta function or a spectral determinant, expanded as a sum over pseudocycles, and ordered by increasing cycle length 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 length $n_p$=$n$ grouped together with pseudocycles whose full itineraries build up the itinerary of $p$. For the binary case this regrouping is given by

$$
\begin{aligned}
1/\zeta &= 1 - t_0 - t_1 - [(t_{01} - t_1 t_0)] - [(t_{001} - t_{01}t_0) + (t_{011} - t_{01}t_1)] \\
&\quad -[(t_{0001} - t_0 t_{001}) + (t_{0111} - t_{011}t_1) \\
&\qquad +(t_{0011} - t_{001}t_1 - t_0 t_{011} + t_0 t_{01} t_1)] - \ldots \\
&= 1 - \sum_f t_f - \sum_n \hat{c}_n \ . \qquad (18.7)
\end{aligned}
$$

All terms in this expansion up to length $n_p = 6$ are given in table **??**. We refer to such regrouped series as *curvature expansions*. .

Such 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 approximants; 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 Markov graph. The terms grouped in brackets are the curvature corrections; the terms grouped in parenthesis are combinations of longer cycles and corresponding sequences of "shadowing" pseudocycles. 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 (13.21). 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 direct averages over periodic orbits such as the trace formulas discussed in sect. 20.5, lies precisely in this organization into nearly canceling combinations: cycle expansions are dominated by short cycles, with long cycles giving exponentially decaying corrections.

[section 13.3]
[section 18.4]

In the case where we know of no finite grammar symbolic dynamics that would help us organize the cycles, the best thing to use is a *stability cutoff* which we shall discuss in sect. 18.5. The idea is to truncate the cycle expansion by including only the pseudocycles 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 $\Lambda_p$ in the data set.

**Table 18.1:** The binary curvature expansion (18.7) up to length 6, listed in such 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.

| | | | |
|---|---|---|---|
| $-t_0$ | | | |
| $-t_1$ | | | |
| $-t_{10}$ | $+t_1t_0$ | | |
| $-t_{100}$ | $+t_{10}t_0$ | | |
| $-t_{101}$ | $+t_{10}t_1$ | | |
| $-t_{1000}$ | $+t_{100}t_0$ | | |
| $-t_{1001}$ | $+t_{100}t_1$ | $+t_{101}t_0$ | $-t_1t_{10}t_0$ |
| $-t_{1011}$ | $+t_{101}t_1$ | | |
| $-t_{10000}$ | $+t_{1000}t_0$ | | |
| $-t_{10001}$ | $+t_{1001}t_0$ | $+t_{1000}t_1$ | $-t_0t_{100}t_1$ |
| $-t_{10010}$ | $+t_{100}t_{10}$ | | |
| $-t_{10101}$ | $+t_{101}t_{10}$ | | |
| $-t_{10011}$ | $+t_{1011}t_0$ | $+t_{1001}t_1$ | $-t_0t_{101}t_1$ |
| $-t_{10111}$ | $+t_{1011}t_1$ | | |
| $-t_{100000}$ | $+t_{10000}t_0$ | | |
| $-t_{100001}$ | $+t_{10001}t_0$ | $+t_{10000}t_1$ | $-t_0t_{1000}t_1$ |
| $-t_{100010}$ | $+t_{10010}t_0$ | $+t_{1000}t_{10}$ | $-t_0t_{100}t_{10}$ |
| $-t_{100011}$ | $+t_{10011}t_0$ | $+t_{10001}t_1$ | $-t_0t_{1001}t_1$ |
| $-t_{100101}$ | $-t_{100110}$ | $+t_{10010}t_1$ | $+t_{10110}t_0$ |
| | $+t_{10}t_{1001}$ | $+t_{100}t_{101}$ | $-t_0t_{10}t_{101}$ $-t_1t_{10}t_{100}$ |
| $-t_{101110}$ | $+t_{10110}t_1$ | $+t_{1011}t_{10}$ | $-t_1t_{101}t_{10}$ |
| $-t_{100111}$ | $+t_{10011}t_1$ | $+t_{10111}t_0$ | $-t_0t_{1011}t_1$ |
| $-t_{101111}$ | $+t_{10111}t_1$ | | |

## 18.2 Construction of cycle expansions

### 18.2.1 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 length $n_p \leq N$ for given fixed $\beta$ and $s$. Denote by subscript $(i)$ the $i$th prime cycle computed, ordered by the topological length $n_{(i)} \leq n_{(i+1)}$. The dynamical zeta function $1/\zeta_N$ truncated to the $n_p \leq N$ cycles is computed recursively, by multiplying

$$1/\zeta_{(i)} = 1/\zeta_{(i-1)}(1 - t_{(i)}z^{n_{(i)}}) , \tag{18.8}$$

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

$$1/\zeta_N = 1 - \sum_{n=1}^{N} c_n z^n . \tag{18.9}$$

In other words, a cycle expansion is a Taylor expansion in the dummy variable $z$ raised to the topological cycle length. If both the number of cycles and their individual weights grow not faster than exponentially with the cycle length, 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 dynamics is given by iterated mapping, the leading zero of (18.9) as 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 (18.9) as function of $s$.

### 18.2.2 Evaluation of traces, spectral determinants

Due to the lack of factorization of the full pseudocycle weight,

$$\det\left(\mathbf{1} - M_{p_1 p_2}\right) \neq \det\left(\mathbf{1} - M_{p_1}\right)\det\left(\mathbf{1} - M_{p_2}\right) ,$$

the cycle expansions for the spectral determinant (17.9) are somewhat less transparent than is the case for the dynamical zeta functions.

We commence the cycle expansion evaluation of a spectral determinant by computing recursively the trace formula (16.10) truncated to all prime cycles $p$ and their repeats such that $n_p r \leq N$:

$$\begin{aligned}
\operatorname{tr}\frac{z\mathcal{L}}{1-z\mathcal{L}}\bigg|_{(i)} &= \operatorname{tr}\frac{z\mathcal{L}}{1-z\mathcal{L}}\bigg|_{(i-1)} + n_{(i)}\sum_{r=1}^{n_{(i)}r\leq N}\frac{e^{(\beta\cdot A_{(i)}-sT_{(i)})r}}{\left|\prod\left(1-\Lambda_{(i),j}^{r}\right)\right|}z^{n_{(i)}r} \\
\operatorname{tr}\frac{z\mathcal{L}}{1-z\mathcal{L}}\bigg|_{N} &= \sum_{n=1}^{N} C_n z^n , \qquad C_n = \operatorname{tr}\mathcal{L}^n . \qquad (18.10)
\end{aligned}$$

This is done numerically: the periodic orbit data set consists of the list of the cycle periods $T_p$, the cycle stability eigenvalues $\Lambda_{p,1}, \Lambda_{p,2}, \ldots, \Lambda_{p,d}$, 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 $(\beta, s)$ parameter values. Now that we have an expansion for the trace formula (16.9) as a power series, we compute the $N$th order approximation to the spectral determinant (17.3),

$$\det(1 - z\mathcal{L})|_N = 1 - \sum_{n=1}^{N} Q_n z^n , \qquad Q_n = n\text{th cumulant} , \qquad (18.11)$$

as follows. The logarithmic derivative relation (17.4) yields

$$\begin{aligned}
\left(\operatorname{tr}\frac{z\mathcal{L}}{1-z\mathcal{L}}\right)\det(1-z\mathcal{L}) &= -z\frac{d}{dz}\det(1-z\mathcal{L}) \\
(C_1 z + C_2 z^2 + \cdots)(1 - Q_1 z - Q_2 z^2 - \cdots) &= Q_1 z + 2Q_2 z^2 + 3Q_3 z^3 \cdots
\end{aligned}$$

so the $n$th order term of the spectral determinant cycle (or in this case, the cumulant) expansion is given recursively by the trace formula expansion coefficients

$$Q_n = \frac{1}{n}\left(C_n - C_{n-1}Q_1 - \cdots C_1 Q_{n-1}\right) , \qquad Q_1 = C_1 . \qquad (18.12)$$

**Table 18.2:** 3-disk repeller escape rates computed from the cycle expansions of the spectral determinant (17.6) and the dynamical zeta function (17.15), as function of the maximal cycle length $N$. The first column indicates the disk-disk center separation to disk radius ratio $R$:$a$, the second column gives the maximal cycle length used, and the third the estimate of the classical escape rate from the fundamental domain spectral determinant cycle expansion. As for larger disk-disk separations the dynamics is more uniform, the convergence is better for $R$:$a = 6$ than for $R$:$a = 3$. For comparison, the fourth column lists a few estimates from from the fundamental domain dynamical zeta function cycle expansion (18.7), and the fifth from the full 3-disk cycle expansion (18.36). 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.)

| $R$:$a$ | $N$ | . $\det(s-\mathcal{A})$ | $1/\zeta(s)$ | $1/\zeta(s)_{\text{3-disk}}$ |
|---|---|---|---|---|
| | 1 | 0.39 | 0.407 | |
| | 2 | 0.4105 | 0.41028 | 0.435 |
| | 3 | 0.410338 | 0.410336 | 0.4049 |
| 6 | 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 | 0.41033840776934648933846130781 92 | | |
| | 1 | 0.41 | | |
| | 2 | 0.72 | | |
| | 3 | 0.675 | | |
| | 4 | 0.67797 | | |
| 3 | 5 | 0.677921 | | |
| | 6 | 0.6779227 | | |
| | 7 | 0.6779226894 | | |
| | 8 | 0.6779226896002 | | |
| | 9 | 0.677922689599532 | | |
| | 10 | 0.67792268959953606 | | |

Given the trace formula (18.10) truncated to $z^N$, we now also have the spectral determinant truncated to $z^N$.

The same program can also be reused to compute the dynamical zeta function cycle expansion (18.9), by replacing $\prod\left(1-\Lambda_{(i),j}^{r}\right)$ in (18.10) by the product of expanding eigenvalues $\Lambda_{(i)} = \prod_e \Lambda_{(i),e}$ (see sect. 17.3).

The calculation of the leading eigenvalue of a given continuous flow evolution operator is now straightforward. After the prime cycles and the pseudocycles have been grouped into subsets of equal topological length, the dummy variable can be set equal to $z = 1$. With $z = 1$, expansion (18.11) is the cycle expansion for (17.6), the spectral determinant $\det(s-\mathcal{A})$ . We vary $s$ in cycle weights, and determine the eigenvalue $s_\alpha$ by finding $s = s_\alpha$ for which (18.11) vanishes. As an example, the convergence of a leading eigenvalue for a nice hyperbolic system is illustrated in table **??** by the listing of pinball escape rate $\gamma$ estimates computed from truncations of (18.7) and (18.11) to different maximal cycle lengths. [chapter 21]

The pleasant surprise is that the coefficients in these cycle expansions can be proven to fall off exponentially or even faster, due to analyticity of $\det(s-\mathcal{A})$ or $1/\zeta(s)$ for $s$ values well beyond those for which the corresponding trace formula diverges. [chapter 21]

### 18.2.3 Newton algorithm for determination of the evolution operator eigenvalues

The cycle expansions of spectral determinants yield the eigenvalues of the evolution operator beyond the leading one. 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 18.1, reveal complicated patterns of resonances even for something so simple as the 3-disk game of pinball. With a good starting guess (such as a location of a zero suggested by the complex $s$ scan of figure 18.1), a zero $1/\zeta(s) = 0$ can now be easily determined by standard numerical methods, such as the iterative Newton algorithm (12.4), with the $m$th Newton estimate given by

$$s_{m+1} = s_m - \left(\zeta(s_m)\frac{\partial}{\partial s}\zeta^{-1}(s_m)\right)^{-1} = s_m - \frac{1/\zeta(s_m)}{\langle T\rangle_\zeta} \ . \tag{18.13}$$

The dominator $\langle T\rangle_\zeta$ required for the Newton iteration is given below, by the cycle expansion (18.22). We need to evaluate it anyhow, as $\langle T\rangle_\zeta$ enters our cycle averaging formulas.

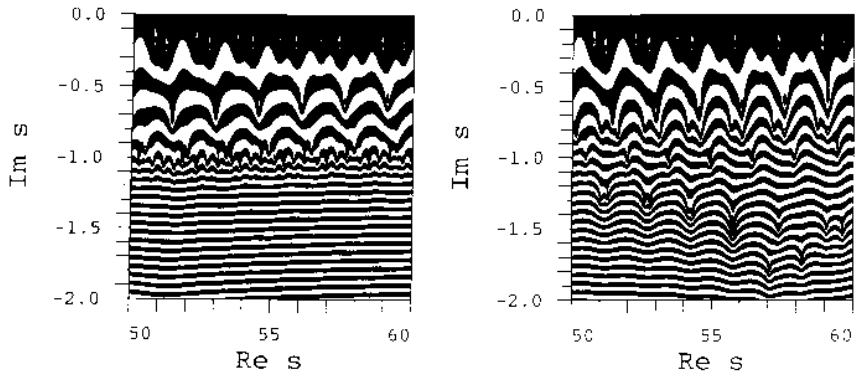



**Figure 18.1:** Examples of the complex $s$ plane scans: contour plots of the logarithm of the absolute values of (a) $1/\zeta(s)$, (b) spectral determinant $\det(s-\mathcal{A})$ for the 3-disk system, separation $a:R=6$, $A_1$ subspace are evaluated numerically. The eigenvalues of the evolution operator $\mathcal{L}$ are given by the centers of elliptic neighborhoods of the rapidly narrowing rings. While the dynamical zeta function is analytic on a strip $\mathrm{Im}\, s \geq -1$, the spectral determinant is entire and reveals further families of zeros. (P.E. Rosenqvist)

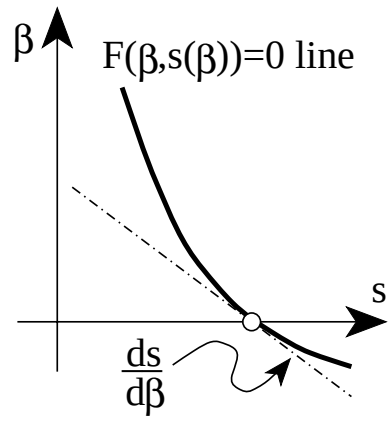



**Figure 18.2:** The eigenvalue condition is satisfied on the curve $F=0$ the $(\beta, s)$ plane. The expectation value of the observable (15.12) is given by the slope of the curve.

## 18.3 Cycle formulas for dynamical averages

The eigenvalue condition in any of the three forms that we have given so far - the level sum (20.18), the dynamical zeta function (18.2), the spectral determinant (18.11):

$$1 = \sum_i^{(n)} t_i\,, \qquad t_i = t_i(\beta, s(\beta))\,, \qquad n_i = n\,, \tag{18.14}$$

$$0 = 1 - \sum_\pi{}' \, t_\pi\,, \qquad t_\pi = t_\pi(z, \beta, s(\beta)) \tag{18.15}$$

$$0 = 1 - \sum_{n=1}^{\infty} Q_n\,, \qquad Q_n = Q_n(\beta, s(\beta))\,, \tag{18.16}$$

is an implicit equation for the eigenvalue $s = s(\beta)$ of form $F(\beta, s(\beta)) = 0$. The eigenvalue $s = s(\beta)$ as a function of $\beta$ is sketched in figure 18.2; the eigenvalue condition is satisfied on the curve $F = 0$. The cycle averaging formulas for the slope and the curvature of $s(\beta)$ are obtained as in (15.12) by taking derivatives of the eigenvalue condition. Evaluated along $F = 0$, the first derivative leads to

$$\begin{aligned} 0 &= \frac{d}{d\beta} F(\beta, s(\beta)) \\ &= \frac{\partial F}{\partial \beta} + \frac{ds}{d\beta}\frac{\partial F}{\partial s}\bigg|_{s=s(\beta)} \quad \Longrightarrow \quad \frac{ds}{d\beta} = -\frac{\partial F}{\partial \beta}\Big/\frac{\partial F}{\partial s}\,, \end{aligned} \tag{18.17}$$

and the second derivative of $F(\beta, s(\beta)) = 0$ yields

$$\frac{d^2 s}{d\beta^2} = -\left[\frac{\partial^2 F}{\partial\beta^2} + 2\frac{ds}{d\beta}\frac{\partial^2 F}{\partial\beta\partial s} + \left(\frac{ds}{d\beta}\right)^2 \frac{\partial^2 F}{\partial s^2}\right] / \frac{\partial F}{\partial s} . \tag{18.18}$$

Denoting by

$$\begin{aligned} \langle A\rangle_F &= -\left.\frac{\partial F}{\partial\beta}\right|_{\beta,s=s(\beta)} , \qquad \langle T\rangle_F = \left.\frac{\partial F}{\partial s}\right|_{\beta,s=s(\beta)} , \\ \left\langle (A-\langle A\rangle)^2\right\rangle_F &= \left.\frac{\partial^2 F}{\partial\beta^2}\right|_{\beta,s=s(\beta)} \end{aligned} \tag{18.19}$$

respectively the mean cycle expectation value of $A$, the mean cycle period, and the second derivative of $F$ computed for $F(\beta, s(\beta)) = 0$, we obtain the cycle averaging formulas for the expectation value of the observable (15.12), and its variance:

$$\langle a\rangle = \frac{\langle A\rangle_F}{\langle T\rangle_F} \tag{18.20}$$

$$\left\langle (a-\langle a\rangle)^2\right\rangle = \frac{1}{\langle T\rangle_F}\left\langle (A-\langle A\rangle)^2\right\rangle_F . \tag{18.21}$$

These formulas are the central result of the periodic orbit theory. As we shall now show, for each choice of the eigenvalue condition function $F(\beta, s)$ in (20.18), (18.2) and (18.11), the above quantities have explicit cycle expansions.

### 18.3.1 Dynamical zeta function cycle expansions

For the dynamical zeta function condition (18.15), the cycle averaging formulas (18.17), (18.21) require evaluation of the derivatives of dynamical zeta function at a given eigenvalue. Substituting the cycle expansion (18.2) for dynamical zeta function we obtain

$$\begin{aligned} \langle A\rangle_\zeta &:= -\frac{\partial}{\partial\beta}\frac{1}{\zeta} = \sum{}' A_\pi t_\pi \\ \langle T\rangle_\zeta &:= \frac{\partial}{\partial s}\frac{1}{\zeta} = \sum{}' T_\pi t_\pi , \quad \langle n\rangle_\zeta := -z\frac{\partial}{\partial z}\frac{1}{\zeta} = \sum{}' n_\pi t_\pi , \end{aligned} \tag{18.22}$$

where the subscript in $\langle\cdots\rangle_\zeta$ stands for the dynamical zeta function average over prime cycles, $A_\pi$, $T_\pi$, and $n_\pi$ are evaluated on pseudocycles (18.4), and pseudocycle 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 $\beta A_\pi - sT_\pi$ in (18.3) vanishes, so the cycle expansions take a simple form

$$\langle A\rangle_\zeta = \sum_\pi{}' (-1)^{k+1} \frac{A_{p_1} + A_{p_2} \cdots + A_{p_k}}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|} , \tag{18.23}$$

and similarly for $\langle T\rangle_\zeta$, $\langle n\rangle_\zeta$. For example, for the complete binary symbolic dynamics the mean cycle period $\langle T\rangle_\zeta$ is given by

$$\begin{aligned}\langle T\rangle_\zeta &= \frac{T_0}{|\Lambda_0|} + \frac{T_1}{|\Lambda_1|} + \left(\frac{T_{01}}{|\Lambda_{01}|} - \frac{T_0 + T_1}{|\Lambda_0\Lambda_1|}\right) \\ &\quad + \left(\frac{T_{001}}{|\Lambda_{001}|} - \frac{T_{01} + T_0}{|\Lambda_{01}\Lambda_0|}\right) + \left(\frac{T_{011}}{|\Lambda_{011}|} - \frac{T_{01} + T_1}{|\Lambda_{01}\Lambda_1|}\right) + \dots .\end{aligned} \tag{18.24}$$

Note that the cycle expansions for averages are grouped into the same shadowing combinations as the dynamical zeta function cycle expansion (18.7), with nearby pseudocycles nearly cancelling each other.

The cycle averaging formulas for the expectation value of the observable $\langle a\rangle$ follow by substitution into (18.21). Assuming zero mean drift $\langle a\rangle = 0$, the cycle expansion (18.11) for the variance $\left\langle (A - \langle A\rangle)^2 \right\rangle_\zeta$ is given by

$$\left\langle A^2 \right\rangle_\zeta = \sum{}' (-1)^{k+1} \frac{\left(A_{p_1} + A_{p_2} \cdots + A_{p_k}\right)^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|} . \tag{18.25}$$

### 18.3.2 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 **??**. For "nice" hyperbolic systems the shadowing ensures exponential convergence of the dynamical zeta function cycle expansions. This, however, is not the best achievable convergence. As has been explained in chapter 21, for such systems the spectral determinant constructed from the same cycle data base is entire, and its cycle expansion converges faster than exponentially. In practice, the best convergence is attained by the spectral determinant cycle expansion (18.16) and its derivatives. The $\partial/\partial s$, $\partial/\partial\beta$ derivatives are in this case computed recursively, by taking derivatives of the spectral determinant cycle expansion contributions (18.12) and (18.10).

The cycle averaging formulas are exact, and highly convergent for nice hyperbolic dynamical systems. An example of its utility is the cycle expansion formula for the Lyapunov exponent of example 18.1. Further applications of cycle expansions will be discussed in chapter 20.

### 18.3.3 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 Poincaré return map. Return times (3.1) might vary wildly, and it is not at all clear that the continuous and discrete time averages are related in any simple way. The relationship turns on to be both elegantly simple, and totally general.

The mean cycle period $\langle T\rangle_\zeta$ 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 $\langle a\rangle$ in terms of continuous time, and would like to also have the corresponding average $\langle a\rangle_{\rm dscr}$ measured in discrete time, given by the number of reflections off billiard walls, the two averages are related by

$$\langle a\rangle_{\rm dscr} = \langle a\rangle \langle T\rangle_\zeta \,/\, \langle n\rangle_\zeta \ , \tag{18.26}$$

where $\langle n\rangle_\zeta$ is the average of the number of bounces $n_p$ along the cycle $p$.

**Example 18.1** ***Cycle expansion formula for Lyapunov exponents:***

*In sect. 15.3 we defined the Lyapunov exponent for a 1-d mapping, related 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 (18.23) yields an exact explict expression for the Lyapunov exponent in terms of prime cycles:*

$$\lambda = \frac{1}{\langle n\rangle_\zeta} \sum{}' (-1)^{k+1} \frac{\log|\Lambda_{p_1}| + \cdots + \log|\Lambda_{p_k}|}{|\Lambda_{p_1}\cdots\Lambda_{p_k}|} \ . \tag{18.27}$$

*For a repeller, the* $1/|\Lambda_p|$ *weights are replaced by normalized measure (20.10)* $\exp(\gamma n_p)/|\Lambda_p|$, *where* $\gamma$ *is the escape rate.*

We mention here without proof that for 2-*d* Hamiltonian flows such as our game of pinball there is only one expanding eigenvalue and (18.27) applies as it stands.

## 18.4 Cycle expansions for finite alphabets

A finite Markov graph like the one given in figure 13.3 (d) is a compact encoding of the transition or the Markov matrix for a given subshift. It is a sparse matrix, and the associated determinant (13.17) can be written down 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:

$$\det(1-T) = 1 - t_0 - t_{0011} - t_{0001} - t_{00011} + t_0 t_{0011} + t_{0011} t_{0001} \tag{18.28}$$

The simplest application of this determinant is to the evaluation of the topological entropy; if we set $t_p = z^{n_p}$, where $n_p$ is the length of the $p$-cycle, the determinant reduces to the topological polynomial (13.18).

The determinant (18.28) is exact for the finite graph figure 13.3 (e), as well as for the associated finite-dimensional transfer operator of example 15.2. For the associated (infinite dimensional) evolution operator, it is the beginning of the cycle expansion of the corresponding dynamical zeta function:

$$\begin{aligned} 1/\zeta &= 1 - t_0 - t_{0011} - t_{0001} + t_{0001}t_{0011} \\ &\qquad -(t_{00011} - t_0 t_{0011} + \ldots \text{curvatures}) \ldots \end{aligned} \tag{18.29}$$

The cycles $\overline{0}$, $\overline{0001}$ and $\overline{0011}$ are the *fundamental* cycles introduced in (18.7); they are not shadowed by any combinations of shorter cycles, and are the basic building blocks of the dynamics.All other cycles appear together with their shadows (for example, the $t_{00011} - t_0 t_{0011}$ combination) and yield exponentially small corrections for hyperbolic systems.

For the cycle counting purposes both $t_{ab}$ and the pseudocycle combination $t_{a+b} = t_a t_b$ in (18.2) have the same weight $z^{n_a+n_b}$, so all curvature combinations $t_{ab} - t_a t_b$ vanish exactly, and the topological polynomial (13.21) offers a quick way of checking the fundamental part of a cycle expansion.

Since for finite grammars the topological zeta functions reduce to polynomials, we are assured that there are just a few fundamental cycles and that all long cycles can be grouped into curvature combinations. For example, the fundamental cycles in exercise 9.2 are the three 2-cycles which bounce back and forth between two disks and the two 3-cycles which visit every disk. 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 lengths of the fundamental cycles are the curvatures $\hat{c}_n$ (in expansion (18.7)), a measure of the deviations between long orbits and their short cycle approximants, expected to fall off rapidly with $n$.

## 18.5 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)

Most dynamical systems of interest have no finite grammar, so at any order in $z$ a cycle expansion may contain unmatched terms which do not fit neatly into the almost cancelling curvature corrections. Similarly, for intermittent systems that we shall discuss in chapter 23, curvature corrections are in general not small, so again the cycle expansions may converge slowly. For such systems schemes

which collect the pseudocycle terms according to some criterion other than the topology of the flow may converge more quickly than expansions based on the topological length.

All chaotic systems exhibit some degree of shadowing, and a good truncation criterion should do its best to respect the shadowing at least approximately. If a long cycle is shadowed by two or more shorter cycles and the flow is smooth, the period and the action will be additive in sense that the period of the longer cycle is approximately the sum of the shorter cycle periods. Similarly, stability is multiplicative, so shadowing is approximately preserved by including all terms with pseudocycle stability

$$\left|\Lambda_{p_1}\cdots\Lambda_{p_k}\right| \leq \Lambda_{\max} \tag{18.30}$$

and ignoring all more unstable pseudocycles.

Two such schemes for ordering cycle expansions which approximately respect shadowing are truncations by the pseudocycle period (or action) and the stability ordering that we shall discuss here. In these schemes a dynamical zeta function or a spectral determinant is expanded keeping all terms for which the period, action or stability for a combination of cycles (pseudocycle) is less than a given cutoff.

The two settings in which the stability ordering may be preferable to the ordering by topological cycle length are the cases of bad grammar and of intermittency.

### 18.5.1 Stability ordering for bad grammars

For generic flows it is often not clear what partition of the 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 (18.7) which 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 length. Another practical advantage of the method (in contrast to Newton method searches) is that it only finds cycles in a given connected ergodic component of state space, ignoring isolated cycles or other ergodic regions elsewhere in the state space.

Why should stability ordered cycle expansion of a dynamical zeta function converge better than the rude trace formula (20.9)? The argument has essentially already been laid out in sect. 13.7: in truncations that respect shadowing most of the pseudocycles appear in shadowing combinations and nearly cancel, while only the relatively small subset affected by the longer and longer pruning rules is not shadowed. So the error is typically of the order of $1/\Lambda$, smaller by factor $e^{hT}$ than

the trace formula (20.9) error, where $h$ is the entropy and $T$ typical cycle length for cycles of stability $\Lambda$.

### 18.5.2 Smoothing

The breaking of exact shadowing cancellations deserves further comment. Partial shadowing which may be present can be (partially) restored by smoothing the stability ordered cycle expansions by replacing the $1/\Lambda$ weight for each term with pseudocycle stability $\Lambda = \Lambda_{p_1} \cdots \Lambda_{p_k}$ by $f(\Lambda)/\Lambda$. Here, $f(\Lambda)$ is a monotonically decreasing function from $f(0) = 1$ to $f(\Lambda_{\max}) = 0$. No smoothing corresponds to a step function.

A typical "shadowing error" induced by the cutoff is due to two pseudocycles of stability $\Lambda$ separated by $\Delta\Lambda$, and whose contribution is of opposite signs. 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 there is an extra term of the form $f'(\Lambda)\Delta\Lambda/\Lambda$, which we want to minimise. A reasonable guess might be to keep $f'(\Lambda)/\Lambda$ constant and as small as possible, that is

$$f(\Lambda) = 1 - \left(\frac{\Lambda}{\Lambda_{\max}}\right)^2$$

The results of a stability ordered expansion (18.30) should always be tested for robustness by varying the cutoff $\Lambda_{\max}$. If this introduces significant variations, smoothing is probably necessary.

### 18.5.3 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 situation 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 that we shall study in detail in chapter 23, the simplest of which is the Farey map

$$f(x) = \begin{cases} f_0 = x/(1-x) & 0 \le x \le 1/2 \\ f_1 = (1-x)/x & 1/2 \le x \le 1 \end{cases}, \tag{18.31}$$

a map which will reappear in the intermittency chapter 23.

For this map the symbolic dynamics is of complete binary type, so lack of shadowing is not due to 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 equals $\Lambda_0 = 1$. This fixed point does not participate directly in the dynamics and is

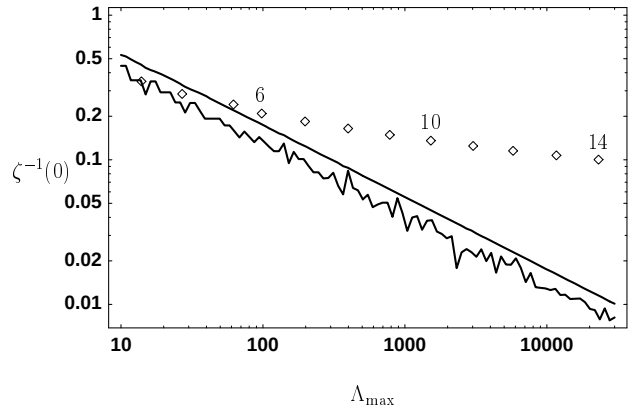



**Figure 18.3:** Comparison of cycle expansion truncation schemes for the Farey map (18.31); the deviation of the truncated cycles expansion for $|1/\zeta_N(0)|$ from the exact flow conservation value $1/\zeta(0) = 0$ is a measure of the accuracy of the truncation. The jagged line is logarithm of the stability ordering truncation error; the smooth line is smoothed according to sect. 18.5.2; the diamonds indicate the error due the topological length truncation, with the maximal cycle length $N$ 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.

omitted from cycle expansions. Its presence is felt in the stabilities of neighboring cycles with $n$ consecutive repeats of the symbol 0's whose stability falls of only as $\Lambda \sim n^2$, in contrast to 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. 13.5.2 leads to a total of 74,248,450 prime cycles of length 30 or less, not including the marginal point $x_0 = 0$. Evaluating a cycle expansion to this order would be no mean computational feat. However, the least unstable cycle omitted has stability of roughly $\Lambda_{10^{30}} \sim 30^2 = 900$, and so amounts to 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-length truncated cycle expansion based on more than $10^9$ pseudocycle terms! On the other hand, truncating by stability at say $\Lambda_{\max} = 3000$, only 409 prime cycles suffice to attain the same accuracy of about 3% error, figure 18.3.

As the Farey map maps the unit interval onto itself, the leading eigenvalue of the Perron-Frobenius operator should equal $s_0 = 0$, so $1/\zeta(0) = 0$. Deviation from this exact result serves as an indication of the convergence of a given cycle expansion. The errors of different truncation schemes are indicated in figure 18.3. We see that topological length 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, the convergence can be improved by going to infinite alphabets.

## 18.6 Dirichlet series

> The most patient reader will thank me for compressing so much nonsense and falsehood into a few lines.
>
> —Gibbon

A Dirichlet series is defined as

$$f(s) = \sum_{j=1}^{\infty} a_j e^{-\lambda_j s} \tag{18.32}$$

where $s$, $a_j$ are complex numbers, and $\{\lambda_j\}$ is a monotonically increasing series of real numbers $\lambda_1 < \lambda_2 < \cdots < \lambda_j < \cdots$. A classical example of a Dirichlet series is the Riemann zeta function for which $a_j = 1$, $\lambda_j = \ln j$. In the present context, formal series over individual pseudocycles such as (18.2) ordered by the increasing pseudocycle periods are often Dirichlet series. For example, for the pseudocycle weight (18.3), the Dirichlet series is obtained by ordering pseudocycles by increasing periods $\lambda_\pi = T_{p_1} + T_{p_2} + \ldots + T_{p_k}$, with the coefficients

$$a_\pi = \frac{e^{\beta\cdot(A_{p_1}+A_{p_2}+\ldots+A_{p_k})}}{\left|\Lambda_{p_1}\Lambda_{p_2}\ldots\Lambda_{p_k}\right|} d_\pi \,,$$

where $d_\pi$ is a degeneracy factor, in the case that $d_\pi$ pseudocycles have the same weight.

If the series $\sum |a_j|$ diverges, the Dirichlet series is absolutely convergent for $\mathrm{Re}\, s > \sigma_{\mathrm{a}}$ and conditionally convergent for $\mathrm{Re}\, s > \sigma_{\mathrm{c}}$, where $\sigma_a$ is the *abscissa of absolute convergence*

$$\sigma_a = \lim_{N\to\infty} \sup \frac{1}{\lambda_N} \ln \sum_{j=1}^{N} |a_j| \,, \tag{18.33}$$

and $\sigma_c$ is the *abscissa of conditional convergence*

$$\sigma_c = \lim_{N\to\infty} \sup \frac{1}{\lambda_N} \ln \left| \sum_{j=1}^{N} a_j \right| \,. \tag{18.34}$$

We shall encounter another example of a Dirichlet series in the semiclassical quantization, the quantum chaos part of `ChaosBook.org`.

## Résumé

A *cycle expansion* is a series representation of a dynamical zeta function, trace formula or a spectral determinant, with products in (17.15) expanded as sums over *pseudocycles*, products of the prime cycle weights $t_p$.

If a flow is hyperbolic and has a topology of a Smale horseshoe (a subshift of finite type), the dynamical zeta functions are holomorphic, 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 barier," 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 that case, the basic observation is that the motion in dynamical systems of few degrees of freedom is in this case 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 approximants; 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 a *difference* of a long cycle $\{ab\}$ minus its shadowing approximation by shorter cycles $\{a\}$ and $\{b\}$:

$$t_{ab} - t_a t_b = t_{ab}(1 - t_a t_b / t_{ab})$$

The orbits that follow the same symbolic dynamics, such as $\{ab\}$ and a "pseudocycle" $\{a\}\{b\}$, lie close to each other, have similar weights, and for longer and longer 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 integrated observable computed, the cycle averaging formulas such as the ones associated with the dynamical zeta function

$$\begin{aligned} \langle a \rangle &= \langle A \rangle_\zeta \,/\, \langle T \rangle_\zeta \\ \langle A \rangle_\zeta &= -\frac{\partial}{\partial \beta}\frac{1}{\zeta} = \sum{}' A_\pi t_\pi \,, \qquad \langle T \rangle_\zeta = \frac{\partial}{\partial s}\frac{1}{\zeta} = \sum{}' T_\pi t_\pi \end{aligned}$$

yield the expectation value (the chaotic, ergodic average over the non–wandering set) of the observable $a(x)$.

## Commentary

**Remark 18.1** Pseudocycle expansions. Bowen's introduction of shadowing $\epsilon$-pseudoorbits [24] was a significant contribution to Smale's theory. Expression "pseudoorbits" seems to have been introduced in the Parry and Pollicott's 1983 paper [4]. Following them M. Berry [9] had used the expression "pseudoorbits" in his 1986 paper on Riemann zeta and quantum chaos. Cycle and curvature expansions of dynamical zeta functions and spectral determinants were introduced in refs. [10, 2]. Some literature [13] refers to the pseudoorbits as

"composite orbits," and to the cycle expansions as "Dirichlet series" (see also remark 18.6 and sect. 18.6).

**Remark 18.2** Cumulant expansion. To a statistical mechanician the curvature expansions are very reminiscent of cumulant expansions. Indeed, (18.12) is the standard Plemelj-Smithies cumulant formula for the Fredholm determinant.The difference is that in cycle expansions each $Q_n$ coefficient is expressed as a sum over exponentially many cycles.

**Remark 18.3** Exponential growth of the number of cycles. Going from $N_n \approx N^n$ periodic points of length $n$ to $M_n$ prime cycles reduces the number of computations from $N_n$ to $M_n \approx N^{n-1}/n$. Use of discrete symmetries (chapter 19) reduces the number of $n$th level terms by another factor. While the reformulation of the theory from the trace (16.28) to the cycle expansion (18.7) thus does not eliminate the exponential growth in the number of cycles, in practice only the shortest cycles are used, and for them the computational labor saving can be significant.

**Remark 18.4** Shadowing cycle-by-cycle. A glance at the low order curvatures in the table **??** leads to the temptation of associating curvatures to individual cycles, such as $\hat{c}_{0001} = t_{0001} - t_0 t_{001}$. 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 [20]; 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 $t_{001}t_{011}$ in table **??** is shared by the two cycles.

**Remark 18.5** Stability ordering. The stability ordering was introduced by Dahlqvist and Russberg [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 [13] for the Lorentz gas which is hyperbolic but the symbolic dynamics is highly pruned, and Dettmann and Cvitanović [14] for a family of intermittent maps. In the applications discussed in the above papers, the stability ordering yields a considerable improvement over the topological length ordering. In quantum chaos applications cycle expansion cancelations are affected by the phases of pseudocycles (their actions), hence *period ordering* rather than stability is frequently employed.

**Remark 18.6** Are cycle expansions Dirichlet series?

Even though some literature [13] refers to cycle expansions as "Dirichlet series," they are not Dirichlet series. Cycle expansions collect contributions of individual cycles into groups that correspond to the coefficients in cumulant expansions of spectral determinants, and the convergence of cycle expansions is controlled by general properties of spectral determinants. Dirichlet series order cycles by their periods or actions, and are only conditionally convergent in regions of interest. The abscissa of absolute convergence is in this context called the "entropy barrier"; contrary to the frequently voiced anxieties, this number does not necessarily has much to do with the actual convergence of the theory.

## Exercises

18.1. **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 problem that follow.

18.2. **Escape rate for a 1-$d$ repeller.** (Continuation of exercise 17.1 - easy, but long)
Consider again the quadratic map (17.31)

$$f(x) = Ax(1-x)$$

on the unit interval, for definitiveness 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 is in the second half), we have a repeller with a complete binary symbolic dynamics.

(a) Sketch the graph of $f$ and determine its two fixed points $\overline{0}$ and $\overline{1}$, together with their stabilities.

(b) Sketch the two branches of $f^{-1}$. Determine all the prime cycles up to topological length 4 using your pocket calculator and backwards iteration of $f$ (see sect. 12.2.1).

(c) Determine the leading zero of the zeta function (17.15) using the weights $t_p = z^{n_p}/|\Lambda_p|$ where $\Lambda_p$ is the stability of the $p$ cycle.

(d) Show that for $A = 9/2$ the escape rate of the repeller is $0.361509\ldots$ using the spectral determinant, with the same cycle weight. If you have taken $A = 6$, the escape rate is in $0.83149298\ldots$, as shown in solution 18.2. Compare the coefficients of the spectral determinant and the zeta function cycle expansions. Which expansion converges faster?

(Per Rosenqvist)

18.3. **Escape rate for the Ulam map.** (Medium; repeat of exercise 12.1) We will try to compute the escape rate for the Ulam map (12.18)

$$f(x) = 4x(1-x),$$

using the method of cycle expansions. The answer should be zero, as nothing escapes.

(a) Compute a few of the stabilities for this map. Show that $\Lambda_0 = 4$, $\Lambda_1 = -2$, $\Lambda_{01} = -4$, $\Lambda_{001} = -8$ and $\Lambda_{011} = 8$.

(b) Show that

$$\Lambda_{\epsilon_1\ldots\epsilon_n} = \pm 2^n$$

and determine a rule for the sign.

(c) (hard) Compute the dynamical zeta function for this system

$$\zeta^{-1} = 1 - t_0 - t_1 - (t_{01} - t_0 t_1) - \cdots$$

You might note that the convergence as function of the truncation cycle length is slow. Try to fix that by treating the $\Lambda_0 = 4$ cycle separately. (Continued as exercise 18.12.)

18.4. **Pinball escape rate, semi-analytical.** Estimate the 3-disk pinball escape rate for $R : a = 6$ by substituting analytical cycle stabilities and periods (exercise 9.3 and exercise 9.4) into the appropriate binary cycle expansion. Compare with the numerical estimate exercise 15.3.

18.5. **Pinball escape rate, from numerical cycles.** Compute the escape rate for $R : a = 6$ 3-disk pinball by substituting list of numerically computed cycle stabilities of exercise 12.5 into the binary cycle expansion.

18.6. **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 (18.5) 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?

18.7. **Counting the 3-disk psudocycles.** (Continuation of exercise 13.12.) Verify that the number of terms in the 3-disk pinball curvature expansion (18.35) is given by

$$\begin{aligned}\prod_p \left(1 + t_p\right) &= \frac{1 - 3z^4 - 2z^6}{1 - 3z^2 - 2z^3} \\ &= 1 + 3z^2 + 2z^3 + \frac{z^4(6 + 12z + 2z^2)}{1 - 3z^2 - 2z^3} \\ &= 1 + 3z^2 + 2z^3 + 6z^4 + 12z^5 \\ &\quad +20z^6 + 48z^7 + 84z^8 + 184z^9 + \ldots\end{aligned}$$

This means that, for example, $c_6$ has a total of 20 terms, in agreement with the explicit 3-disk cycle expansion (18.36).

18.8. **3–disk unfactorized zeta cycle expansions.** Check that the curvature expansion (18.2) for the 3-disk pinball, assuming no symmetries between disks, is given by

$$\begin{aligned}
1/\zeta &= (1-z^2t_{12})(1-z^2t_{13})(1-z^2t_{23}) \\
&\quad (1-z^3t_{123})(1-z^3t_{132})(1-z^4t_{1213}) \\
&\quad (1-z^4t_{1232})(1-z^4t_{1323})(1-z^5t_{12123})\cdots \\
&= 1-z^2t_{12}-z^2t_{23}-z^2t_{31}-z^3(t_{123}+t_{132}) \\
&\quad -z^4[(t_{1213}-t_{12}t_{13})+(t_{1232}-t_{12}t_{23}) \\
&\quad +(t_{1323}-t_{13}t_{23})] \\
&\quad -z^5[(t_{12123}-t_{12}t_{123})+\cdots]-\cdots \qquad (18.35)
\end{aligned}$$

The symmetrically arranged 3-disk pinball cycle expansion of the Euler product (18.2) (see table **??** and figure 9.3) is given by:

$$\begin{aligned}
1/\zeta &= (1-z^2t_{12})^3(1-z^3t_{123})^2(1-z^4t_{1213})^3 \\
&\quad (1-z^5t_{12123})^6(1-z^6t_{121213})^6 \\
&\quad (1-z^6t_{121323})^3\ldots \\
&= 1-3z^2\,t_{12}-2z^3\,t_{123}-3z^4\,(t_{1213}-t_{12}^2) \\
&\quad -6z^5\,(t_{12123}-t_{12}t_{123}) \\
&\quad -z^6\,(6\,t_{121213}+3\,t_{121323}+t_{12}^3-9\,t_{12}t_{1213}-t_{123}^2) \\
&\quad -6z^7\,(t_{1212123}+t_{1212313}+t_{1213123}+t_{12}^2t_{123} \\
&\quad -3\,t_{12}t_{12123}-t_{123}t_{1213}) \\
&\quad -3z^8\,(2\,t_{12121213}+t_{12121313}+2\,t_{12121323} \\
&\quad +2\,t_{12123123}+2\,t_{12123213}+t_{12132123} \\
&\qquad +3\,t_{12}^2t_{1213}+t_{12}t_{123}^2-6\,t_{12}t_{121213} \\
&\qquad -3\,t_{12}t_{121323}-4\,t_{123}t_{12123}-t_{1213}^2)-\cdots \qquad (18.36)
\end{aligned}$$

**Remark 18.7** Unsymmetrized cycle expansions. The above 3-disk cycle expansions might be useful for cross-checking purposes, but, as we shall see in chapter 19, they are not recommended for actual computations, as the factorized zeta functions yield much better convergence.

18.9. **4–disk unfactorized dynamical zeta function cycle expansions** For the symmetriclly arranged 4-disk pinball the symmetry group is $C_{4v}$, of order 8. The degenerate cycles can have multiplicities 2, 4 or 8 (see table **??**):

$$\begin{aligned}
1/\zeta &= (1-z^2t_{12})^4(1-z^2t_{13})^2(1-z^3t_{123})^8 \\
&\quad (1-z^4t_{1213})^8(1-z^4t_{1214})^4(1-z^4t_{1234})^2 \\
&\quad (1-z^4t_{1243})^4(1-z^5t_{12123})^8(1-z^5t_{12124})^8 \\
&\quad (1-z^5t_{12134})^8(1-z^5t_{12143})^8 \\
&\quad (1-z^5t_{12313})^8(1-z^5t_{12413})^8\cdots \qquad (18.37)
\end{aligned}$$

and the cycle expansion is given by

$$\begin{aligned}
1/\zeta &= 1-z^2(4\,t_{12}+2\,t_{13})-8z^3\,t_{123} \\
&\quad -z^4(8\,t_{1213}+4\,t_{1214}+2\,t_{1234}+4\,t_{1243} \\
&\quad -6\,t_{12}^2-t_{13}^2-8\,t_{12}t_{13}) \\
&\quad -8z^5(t_{12123}+t_{12124}+t_{12134}+t_{12143}+t_{12313} \\
&\quad +t_{12413}-4\,t_{12}t_{123}-2\,t_{13}t_{123}) \\
&\quad -4z^6(2\,S_8+S_4+t_{12}^3+3\,t_{12}^2\,t_{13}+t_{12}t_{13}^2 \\
&\quad -8\,t_{12}t_{1213}-4\,t_{12}t_{1214} \\
&\quad -2\,t_{12}t_{1234}-4\,t_{12}t_{1243} \\
&\quad -4\,t_{13}t_{1213}-2\,t_{13}t_{1214}-t_{13}t_{1234} \\
&\quad -2\,t_{13}t_{1243}-7\,t_{123}^2)-\cdots \qquad (18.38)
\end{aligned}$$

where in the coefficient to $z^6$ 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 of multiplicity 4, respectively; the orbits are listed in table **??**.

18.10. **Tail resummations.** A simple illustration of such tail resummation is the $\zeta$ function for the Ulam map (12.18) for which the cycle structure is exceptionally simple: the eigenvalue of the $x_0 = 0$ fixed point is 4, while the eigenvalue of any other $n$-cycle is $\pm 2^n$. Typical cycle weights used in thermodynamic averaging are $t_0 = 4^\tau z$, $t_1 = t = 2^\tau z$, $t_p = t^{n_p}$ for $p \neq 0$. The simplicity of the cycle eigenvalues enables us to evaluate the $\zeta$ function by a simple trick: we note that if the value of any $n$-cycle eigenvalue were $t^n$, (17.21) would yield $1/\zeta = 1-2t$. There is only one cycle, the $x_0$ fixed point, that has a different weight $(1-t_0)$, so we factor it out, multiply the rest by $(1-t)/(1-t)$, and obtain a rational $\zeta$ function

$$1/\zeta(z) = \frac{(1-2t)(1-t_0)}{(1-t)} \qquad (18.39)$$

Consider how we would have detected the pole at $z = 1/t$ without the above trick. As the $\overline{0}$ fixed point is isolated in its stability, we would have kept the factor $(1-t_0)$ in (18.7) unexpanded, and noted that all curvature combinations in (18.7) which include the $t_0$ factor are unbalanced, so that the cycle expansion is an infinite series:

$$\prod_p \left(1-t_p\right) = (1-t_0)(1-t-t^2-t^3-t^4-\ldots) \qquad (18.40)$$

(we shall return to such infinite series in chapter 23). The geometric series in the brackets sums up to (18.39). Had we expanded the $(1-t_0)$ factor, we would have noted that the ratio of the successive curvatures is exactly $c_{n+1}/c_n = t$; summing we would recover the rational $\zeta$ function (18.39).

18.11. **Escape rate for the Rössler flow.** (continuation of exercise 12.7) Try to compute the escape rate for the Rössler flow (2.17) 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 12.7.

18.12. **Ulam map is conjugate to the tent map.** (Continuation of exercise 18.3 / repeat of exercise 6.3 and exercise 12.2; requires real smarts, unless you look it up) Explain the magically simple form of cycle stabilities of exercise 18.3 by constructing an explicit smooth conjugacy (6.1)

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

that conjugates the Ulam map (12.18) into the tent map (10.6).

## References


[18.1] P. Cvitanović, *Phys. Rev. Lett.* **61**, 2729 (1988).

[18.2] R. Artuso, E. Aurell and P. Cvitanović, "Recycling of strange sets I: Cycle expansions," *Nonlinearity* **3**, 325 (1990).

[18.3] R. Artuso, E. Aurell and P. Cvitanović, "Recycling of strange sets II: Applications," *Nonlinearity* **3**, 361 (1990).

[18.4] S. Grossmann and S. Thomae, *Z. Naturforsch.* **32 a**, 1353 (1977); reprinted in ref. [5].

[18.5] *Universality in Chaos*, P. Cvitanović, ed., (Adam Hilger, Bristol 1989).

[18.6] F. Christiansen, P. Cvitanović and H.H. Rugh, *J. Phys* **A 23**, L713 (1990).

[18.7] J. Plemelj, "Zur Theorie der Fredholmschen Funktionalgleichung," *Monat. Math. Phys.* **15**, 93 (1909).

[18.8] F. Smithies, "The Fredholm theory of integral equations," *Duke Math.* **8**, 107 (1941).

[18.9] M.V. Berry, in *Quantum Chaos and Statistical Nuclear Physics*, ed. T.H. Seligman and H. Nishioka, *Lecture Notes in Physics* **263**, 1 (Springer, Berlin, 1986).

[18.10] P. Cvitanović, "Invariant measurements of strange sets in terms of cycles," *Phys. Rev. Lett.* **61**, 2729 (1988).

[18.11] B. Eckhardt and G. Russberg, *Phys. Rev.* **E 47**, 1578 (1993).

[18.12] P. Dahlqvist and G. Russberg, "Periodic orbit quantization of bound chaotic systems," *J. Phys.* **A 24**, 4763 (1991); P. Dahlqvist *J. Phys.* **A 27**, 763 (1994).

[18.13] C. P. Dettmann and G. P. Morriss, *Phys. Rev. Lett.* **78**, 4201 (1997).

[18.14] C. P. Dettmann and P. Cvitanović, *Cycle expansions for intermittent diffusion Phys. Rev. E* **56**, 6687 (1997); chao-dyn/9708011.

# Chapter 19

# Discrete factorization

> No endeavor that is worthwhile is simple in prospect; if it is right, it will be simple in retrospect.
>
> —Edward Teller

THE UTILITY of discrete symmetries in reducing spectrum calculations is familiar from quantum mechanics. Here we show that the classical spectral determinants factor in essentially the same way as the quantum ones. In the process we 1.) learn that the classical 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/PDEs, 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 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 $m_p$ to a dynamical zeta function factorizes into a product over the $d_\alpha$-dimensional irreducible representations $D_\alpha$ of the symmetry group,

$$(1-t_p)^{m_p} = \prod_\alpha \det\left(1 - D_\alpha(h_{\tilde{p}})t_{\tilde{p}}\right)^{d_\alpha}, \quad t_p = t_{\tilde{p}}^{g/m_p},$$

where $t_{\tilde{p}}$ is the cycle weight evaluated on the relative periodic orbit $\tilde{p}$, $g = |G|$ is the order of the group, $h_{\tilde{p}}$ is the group element relating the fundamental domain cycle $\tilde{p}$ 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. $p$-cycle into a cycle weight $t_p$.

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 the definitions relegated to appendix H.1. We develop here the cycle expansions for factorized determinants, and exemplify them by working two cases of physical interest: $C_2 = D_1$, $C_{3v} = D_3$ symmetries. $C_{2v} = D_2 \times D_2$ and $C_{4v} = D_4$ symmetries are discussed in appendix H.

## 19.1 Preview

As we saw in chapter 9, 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 classical 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.

### 19.1.1 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 9.1. Denote the reflection operation by $Rx = -x$. The symmetry of the map implies that if $\{x_n\}$ is a trajectory, than also $\{Rx_n\}$ is a trajectory because $Rx_{n+1} = Rf(x_n) = f(Rx_n)$. 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 $R$. Furthermore, the evolution operator commutes with $R$, $\mathcal{L}(y,x) = \mathcal{L}(Ry,Rx)$. $R$ satisfies $R^2 = \mathbf{e}$ and can be used to decompose the state space into mutually orthogonal symmetric and antisymmetric subspaces by means of projection operators

$$
\begin{aligned}
P_{A_1} &= \frac{1}{2}(\mathbf{e}+R)\,, \qquad P_{A_2} = \frac{1}{2}(\mathbf{e}-R)\,, \\
\mathcal{L}_{A_1}(y,x) &= P_{A_1}\mathcal{L}(y,x) = \frac{1}{2}\left(\mathcal{L}(y,x)+\mathcal{L}(-y,x)\right)\,, \\
\mathcal{L}_{A_2}(y,x) &= P_{A_2}\mathcal{L}(y,x) = \frac{1}{2}\left(\mathcal{L}(y,x)-\mathcal{L}(-y,x)\right)\,. \qquad (19.1)
\end{aligned}
$$

To compute the traces of the symmetrization and antisymmetrization projection operators (19.1), 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 $\{x_a\} \cap \{Rx_a\} = \emptyset$, where $\{x_a\}$ is the set of periodic points belonging to the cycle $a$. Thus $R$ 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 (19.1); as they are degenerate, the prefactor $1/2$ cancels. Resuming as in the derivation of (17.15) we find that asymmetric orbits yield the same contribution to the symmetric and the antisymmetric subspaces:

$$\det(1-\mathcal{L}_\pm)_a = \prod_a \prod_{k=0}^{\infty}\left(1-\frac{t_a}{\Lambda_a^k}\right), \quad t_a = \frac{z^{n_a}}{|\Lambda_a|}.$$

**Symmetric cycles:** A cycle $s$ is reflection symmetric if operating with $R$ on the set of cycle points reproduces the set. The period of a symmetric cycle is always even ($n_s = 2n_{\tilde{s}}$) and the mirror image of the $x_s$ cycle point is reached by traversing the irreducible segment $\tilde{s}$ of length $n_{\tilde{s}}$, $f^{n_{\tilde{s}}}(x_s) = Rx_s$. $\delta(x - f^n(x))$ 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 stability eigenvalue 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 $\mathcal{M}_{\tilde{s}}$ of the $s$ cycle, we obtain the contribution to $\mathrm{tr}\,\mathcal{L}_\pm^n$:

$$\begin{aligned} z^n \mathrm{tr}\,\mathcal{L}_\pm^n &\to \int_{\mathcal{M}_s} dx\, z^n \frac{1}{2}\left(\delta(x-f^n(x)) \pm \delta(x+f^n(x))\right) \\ &= n_{\tilde{s}}\left(\sum_{r=2}^{\text{even}} \delta_{n,rn_{\tilde{s}}} \frac{t_{\tilde{s}}^r}{1-1/\Lambda_{\tilde{s}}^r} \pm \sum_{r=1}^{\text{odd}} \delta_{n,rn_{\tilde{s}}} \frac{t_{\tilde{s}}^r}{1-1/\Lambda_{\tilde{s}}^r}\right) \\ &= n_{\tilde{s}} \sum_{r=1}^{\infty} \delta_{n,rn_{\tilde{s}}} \frac{(\pm t_{\tilde{s}})^r}{1-1/\Lambda_{\tilde{s}}^r}. \end{aligned}$$

Substituting all symmetric cycles $s$ into $\det(1-\mathcal{L}_\pm)$ and resuming we obtain:

$$\det(1-\mathcal{L}_\pm)_{\tilde{s}} = \prod_{\tilde{s}} \prod_{k=0}^{\infty}\left(1 \mp \frac{t_{\tilde{s}}}{\Lambda_{\tilde{s}}^k}\right)$$

**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 $\delta(x + f^n(x))$:

$$
\begin{aligned}
z^n \mathrm{tr}\, \mathcal{L}_\pm^n &\to \int_{\mathcal{M}_b} dx\, z^n \frac{1}{2} \left(\delta(x - f^n(x)) \pm \delta(x + f^n(x))\right) \\
&= \sum_{r=1}^{\infty} \delta_{n,r}\, t_b^r \frac{1}{2} \left( \frac{1}{1 - 1/\Lambda_b^r} \pm \frac{1}{1 + 1/\Lambda_b^r} \right) \\
z^n \,\mathrm{tr}\, \mathcal{L}_+^n &\to \sum_{r=1}^{\infty} \delta_{n,r} \frac{t_b^r}{1 - 1/\Lambda_b^{2r}} \,; \qquad z^n \,\mathrm{tr}\, \mathcal{L}_-^n \to \sum_{r=1}^{\infty} \delta_{n,r} \frac{1}{\Lambda_b^r} \frac{t_b^r}{1 - 1/\Lambda_b^{2r}} \,.
\end{aligned}
$$

Boundary orbit contributions to the factorized spectral determinants follow by resummation:

$$
\det(1 - \mathcal{L}_+)_b = \prod_{k=0}^{\infty} \left( 1 - \frac{t_b}{\Lambda_b^{2k}} \right) , \qquad \det(1 - \mathcal{L}_-)_b = \prod_{k=0}^{\infty} \left( 1 - \frac{t_b}{\Lambda_b^{2k+1}} \right)
$$

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:

$$
F_+(z) = \prod_a \prod_{k=0}^{\infty} \left( 1 - \frac{t_a}{\Lambda_a^k} \right) \prod_{\tilde{s}} \prod_{k=0}^{\infty} \left( 1 - \frac{t_{\tilde{s}}}{\Lambda_{\tilde{s}}^k} \right) \prod_{k=0}^{\infty} \left( 1 - \frac{t_b}{\Lambda_b^{2k}} \right)
$$

$$
F_-(z) = \prod_a \prod_{k=0}^{\infty} \left( 1 - \frac{t_a}{\Lambda_a^k} \right) \prod_{\tilde{s}} \prod_{k=0}^{\infty} \left( 1 + \frac{t_{\tilde{s}}}{\Lambda_{\tilde{s}}^k} \right) \prod_{k=0}^{\infty} \left( 1 - \frac{t_b}{\Lambda_b^{2k+1}} \right) \tag{19.2}
$$

We shall work out the symbolic dynamics of such reflection symmetric systems in some detail in sect. 19.5. 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-$d$ maps. We now turn to discussion of the general case.

[exercise 19.1]

## 19.2 Discrete symmetries

A dynamical system is invariant under a symmetry group $G = \{e, g_2, \ldots, g_{|G|}\}$ if the equations of motion are invariant under all symmetries $g \in G$. For a map $x_{n+1} = f(x_n)$ and the evolution operator $\mathcal{L}(y, x)$ defined by (15.23) this means

$$
\begin{aligned}
f(x) &= \mathbf{g}^{-1} f(\mathbf{g}x) \\
\mathcal{L}(y, x) &= \mathcal{L}(\mathbf{g}y, \mathbf{g}x) \,.
\end{aligned} \tag{19.3}
$$

Bold face letters for group elements indicate a suitable representation on state space. For example, if a 2-dimensional map has the symmetry $x_1 \to -x_1$, $x_2 \to -x_2$, the symmetry group $G$ consists of the identity and $C$, a rotation by $\pi$ around the origin. The map $f$ must then commute with rotations by $\pi$, $f(Rx) = \mathbf{C}f(x)$, with $R$ given by the $[2 \times 2]$ matrix

$$R = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{19.4}$$

$R$ satisfies $R^2 = e$ and can be used to decompose the state space into mutually orthogonal symmetric and antisymmetric subspaces by means of projection operators (19.1). More generally the projection operator onto the $\alpha$ irreducible subspace of dimension $d_\alpha$ is given by $P_\alpha = (d_\alpha/|G|)\sum \chi_\alpha(h)\mathbf{h}^{-1}$, where $\chi_\alpha(h) = \mathrm{tr}\, D_\alpha(h)$ are the group characters, and the transfer operator $\mathcal{L}$ splits into a sum of inequivalent irreducible subspace contributions $\sum_\alpha \mathrm{tr}\, \mathcal{L}_\alpha$,

$$\mathcal{L}_\alpha(y, x) = \frac{d_\alpha}{|G|} \sum_{h \in G} \chi_\alpha(h) \mathcal{L}(\mathbf{h}^{-1}y, x) \ . \tag{19.5}$$

The prefactor $d_\alpha$ in the above reflects the fact that a $d_\alpha$-dimensional representation occurs $d_\alpha$ times.

### 19.2.1 Cycle degeneracies

Taking into account these degeneracies, the Euler product (17.15) takes the form

$$\prod_p (1 - t_p) = \prod_{\hat{p}} (1 - t_{\hat{p}})^{m_{\hat{p}}} . \tag{19.6}$$

The Euler product (17.15) for the $C_{3v}$ symmetric 3-disk problem is given in (18.36).

## 19.3 Dynamics in the fundamental domain

If the dynamics is invariant under a discrete symmetry, the state space $M$ can be completely tiled by the fundamental domain $\tilde{M}$ and its images $a\tilde{M}$, $b\tilde{M}$, ... under the action of the symmetry group $G = \{e, a, b, \ldots\}$,

$$M = \sum_{a \in G} M_a = \sum_{a \in G} a\tilde{M} .$$

In the above example (19.4) with symmetry group $G = \{e, C\}$, the state space $M = \{x_1\text{-}x_2 \text{ plane}\}$ can be tiled by a fundamental domain $\tilde{M} = \{\text{half-plane } x_1 \geq 0\}$, and $\mathbf{C}\tilde{M} = \{\text{half-plane } x_1 \leq 0\}$, its image under rotation by $\pi$.

If the space $M$ is decomposed into $g$ tiles, a function $\phi(x)$ over $M$ splits into a $g$-dimensional vector $\phi_a(x)$ defined by $\phi_a(x) = \phi(x)$ if $x \in M_a$, $\phi_a(x) = 0$ otherwise. Let $h = ab^{-1}$ conflicts with be the symmetry operation that maps the endpoint domain $M_b$ into the starting point domain $M_a$, and let $D(h)_{ba}$, the left regular representation, be the $[g \times g]$ matrix whose $b, a$-th entry equals unity if $a = hb$ and zero otherwise; $D(h)_{ba} = \delta_{bh,a}$. Since the symmetries act on state space as well, the operation $h$ enters in two guises: as a $[g \times g]$ matrix $D(h)$ which simply permutes the domain labels, and as a $[d \times d]$ matrix representation $\mathbf{h}$ of a discrete symmetry operation on the $d$ state space coordinates. For instance, in the above example (19.4) $h \in C_2$ and $D(h)$ can be either the identity or the interchange of the two domain labels,

$$D(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} , \quad D(C) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} . \tag{19.7}$$

Note that $D(h)$ is a permutation matrix, mapping a tile $M_a$ into a different tile $M_{ha} \neq M_a$ if $h \neq e$. Consequently only $D(e)$ has diagonal elements, and $\mathrm{tr}\, D(h) = g\delta_{h,e}$. However, the state space transformation $\mathbf{h} \neq \mathbf{e}$ leaves invariant sets of *boundary* points; for example, under reflection $\sigma$ across a symmetry axis, the axis itself remains invariant. The boundary periodic orbits that belong to such pointwise invariant sets will require special care in $\mathrm{tr}\, \mathcal{L}$ evaluations.

One can associate to the evolution operator (15.23) a $[g \times g]$ matrix evolution operator defined by

$$\mathcal{L}_{ba}(y, x) = D(h)_{ba} \mathcal{L}(y, x) ,$$

if $x \in M_a$ and $y \in M_b$, and zero otherwise. Now we can use the invariance condition (19.3) to move the starting point $x$ into the fundamental domain $x = \mathbf{a}\tilde{x}$, $\mathcal{L}(y, x) = \mathcal{L}(\mathbf{a}^{-1}y, \tilde{x})$, and then use the relation $a^{-1}b = h^{-1}$ to also relate the endpoint $y$ to its image in the fundamental domain, $\tilde{\mathcal{L}}(\tilde{y}, \tilde{x}) := \mathcal{L}(\mathbf{h}^{-1}\tilde{y}, \tilde{x})$. With this operator which is restricted to the fundamental domain, the global dynamics reduces to

$$\mathcal{L}_{ba}(y, x) = D(h)_{ba} \tilde{\mathcal{L}}(\tilde{y}, \tilde{x}) .$$

While the global trajectory runs over the full space $M$, the restricted trajectory is brought back into the fundamental domain $\tilde{M}$ any time it crosses into adjoining tiles; the two trajectories are related by the symmetry operation $h$ which maps the global endpoint into its fundamental domain image.

Now the traces (17.3) required for the evaluation of the eigenvalues of the transfer operator can be evaluated on the fundamental domain alone

$$\mathrm{tr}\, \mathcal{L} = \int_M dx \mathcal{L}(x, x) = \int_{\tilde{M}} d\tilde{x} \sum_h \mathrm{tr}\, D(h) \; \mathcal{L}(\mathbf{h}^{-1}\tilde{x}, \tilde{x}) \tag{19.8}$$

The fundamental domain integral $\int d\tilde{x}\ \mathcal{L}(\mathbf{h}^{-1}\tilde{x},\tilde{x})$ picks up a contribution from every global cycle (for which $h = e$), but it also picks up contributions from shorter segments of global cycles. The permutation matrix $D(h)$ guarantees by the identity $\mathrm{tr}\, D(h) = 0$, $h \neq e$, that only those repeats of the fundamental domain cycles $\tilde{p}$ that correspond to complete global cycles $p$ contribute. Compare, for example, the contributions of the $\overline{12}$ and $\overline{0}$ cycles of figure 11.2. $\mathrm{tr}\, D(h)\tilde{\mathcal{L}}$ does not get a contribution from the $\overline{0}$ cycle, as the symmetry operation that maps the first half of the $\overline{12}$ into the fundamental domain is a reflection, and $\mathrm{tr}\, D(\sigma) = 0$. In contrast, $\sigma^2 = e$, $\mathrm{tr}\, D(\sigma^2) = 6$ insures that the repeat of the fundamental domain fixed point $\mathrm{tr}\,(D(h)\tilde{\mathcal{L}})^2 = 6t_0^2$, gives the correct contribution to the global trace $\mathrm{tr}\,\mathcal{L}^2 = 3 \cdot 2t_{12}$.

Let $p$ be the full orbit, $\tilde{p}$ the orbit in the fundamental domain and $h_{\tilde{p}}$ an element of $\mathcal{H}_p$, the symmetry group of $p$. Restricting the volume integrations to the infinitesimal neighborhoods of the cycles $p$ and $\tilde{p}$, respectively, and performing the standard resummations, we obtain the identity

$$(1 - t_p)^{m_p} = \det\left(1 - D(h_{\tilde{p}})t_{\tilde{p}}\right) \ , \tag{19.9}$$

valid cycle by cycle in the Euler products (17.15) for $\det(1-\mathcal{L})$. Here "det" refers to the $[g \times g]$ matrix representation $D(h_{\tilde{p}})$; as we shall see, this determinant can be evaluated in terms of standard characters, and no explicit representation of $D(h_{\tilde{p}})$ is needed. Finally, if a cycle $p$ is invariant under the symmetry subgroup $\mathcal{H}_p \subseteq G$ of order $h_p$, its weight can be written as a repetition of a fundamental domain cycle

$$t_p = t_{\tilde{p}}^{h_p} \tag{19.10}$$

computed on the irreducible segment that corresponds to a fundamental domain cycle. For example, in figure 11.2 we see by inspection that $t_{12} = t_0^2$ and $t_{123} = t_1^3$.

### 19.3.1 Boundary orbits

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 fundamental 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 $1/\det(\mathbf{1}-\mathbf{h}M_p)$, where $\mathbf{h}$ is a symmetry operation that leaves the orbit pointwise invariant; see for example sect. 19.1.1.

Consequences for the dynamical zeta function factorizations are that sometimes a boundary orbit does not contribute. A derivation of a dynamical zeta function (17.15) from a determinant like (17.9) 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 (17.15). 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 surface of 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 $\mathcal{H}_p = \{e, b_2, \ldots, b_h\}$, then the corresponding group element in (19.9) will be replaced by a projector. If the weights are insensitive to the signs of the eigenvalues, then this projector is

$$g_p = \frac{1}{h}\sum_{i=1}^{h} b_i \,. \tag{19.11}$$

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 (19.11) is replaced by

$$g_p = \frac{1}{h}\sum_{i=1}^{h} \epsilon(b_i)\, b_i \,. \tag{19.12}$$

We have illustrated the above in sect. 19.1.1 by working out the full factorization for the 1-dimensional reflection symmetric maps.

## 19.4 Factorizations of dynamical zeta functions

In chapter 9 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.

### 19.4.1 Factorizations of dynamical dynamical zeta functions

According to (19.9) and (19.10), the contribution of a degenerate class of global cycles (cycle $p$ with multiplicity $m_p = g/h_p$) to a dynamical zeta function is given by the corresponding fundamental domain cycle $\tilde{p}$:

$$(1 - t_{\tilde{p}}^{h_p})^{g/h_p} = \det\left(1 - D(h_{\tilde{p}})t_{\tilde{p}}\right) \tag{19.13}$$

Let $D(h) = \bigoplus_\alpha d_\alpha D_\alpha(h)$ be the decomposition of the matrix representation $D(h)$ into the $d_\alpha$ dimensional irreducible representations $\alpha$ of a finite group $G$. Such decompositions are block-diagonal, so the corresponding contribution to the Euler product (17.9) factorizes as

$$\det(1 - D(h)t) = \prod_\alpha \det(1 - D_\alpha(h)t)^{d_\alpha} \ , \tag{19.14}$$

where now the product extends over all distinct $d_\alpha$-dimensional irreducible representations, each contributing $d_\alpha$ 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 (19.13); but from the transfer operator point of view, the key observation is that the symmetry reduces the transfer operator to a block diagonal form; this block diagonalization implies that the dynamical zeta functions (17.15) factorize as

$$\frac{1}{\zeta} = \prod_\alpha \frac{1}{\zeta_\alpha^{d_\alpha}} \ , \qquad \frac{1}{\zeta_\alpha} = \prod_{\tilde{p}} \det\left(1 - D_\alpha(h_{\tilde{p}})t_{\tilde{p}}\right) \ . \tag{19.15}$$

Determinants of $d$-dimensional irreducible representations can be evaluated using the expansion of determinants in terms of traces,

$$\begin{aligned}\det(1 + M) \quad = \quad & 1 + \operatorname{tr} M + \frac{1}{2}\left((\operatorname{tr} M)^2 - \operatorname{tr} M^2\right) \\ & + \frac{1}{6}\left((\operatorname{tr} M)^3 - 3(\operatorname{tr} M)(\operatorname{tr} M^2) + 2\operatorname{tr} M^3\right) \\ & + \cdots + \frac{1}{d!}\left((\operatorname{tr} M)^d - \cdots\right) \ ,\end{aligned} \tag{19.16}$$

and each factor in (19.14) can be evaluated by looking up the characters $\chi_\alpha(h) = \operatorname{tr} D_\alpha(h)$ in standard tables [10]. In terms of characters, we have for the 1-dimensional representations

$$\det(1 - D_\alpha(h)t) = 1 - \chi_\alpha(h)t \ ,$$

for the 2-dimensional representations

$$\det(1 - D_\alpha(h)t) = 1 - \chi_\alpha(h)t + \frac{1}{2}\left(\chi_\alpha(h)^2 - \chi_\alpha(h^2)\right)t^2,$$

and so forth.

In the fully symmetric subspace $\mathrm{tr}\, D_{A_1}(h) = 1$ 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 transfer operator. [exercise 19.2]

### 19.4.2 Factorizations of spectral determinants

Factorization of the full spectral determinant (17.3) proceeds in essentially the same manner as the factorization of dynamical zeta functions outlined above. By (19.5) and (19.8) the trace of the transfer operator $\mathcal{L}$ splits into the sum of inequivalent irreducible subspace contributions $\sum_\alpha \mathrm{tr}\, \mathcal{L}_\alpha$, with

$$\mathrm{tr}\, \mathcal{L}_\alpha = d_\alpha \sum_{h\in G} \chi_\alpha(h) \int_{\tilde{M}} d\tilde{x}\, \mathcal{L}(\mathbf{h}^{-1}\tilde{x}, \tilde{x})\,.$$

This leads by standard manipulations to the factorization of (17.9) into

$$\begin{aligned} F(z) &= \prod_\alpha F_\alpha(z)^{d_\alpha} \\ F_\alpha(z) &= \exp\left(-\sum_{\tilde{p}} \sum_{r=1}^{\infty} \frac{1}{r} \frac{\chi_\alpha(h_{\tilde{p}}^r) z^{n_{\tilde{p}} r}}{|\det\left(\mathbf{1} - \tilde{M}_{\tilde{p}}^r\right)|}\right), \end{aligned} \tag{19.17}$$

where $\tilde{M}_{\tilde{p}} = \mathbf{h}_{\tilde{p}} M_{\tilde{p}}$ is the fundamental domain Jacobian. Boundary orbits require special treatment, discussed in sect. 19.3.1, with examples given in the next section as well as in the specific factorizations discussed below.

The factorizations (19.15), (19.17) are the central formulas of this chapter. We now work out the group theory factorizations of cycle expansions of dynamical zeta functions for the cases of $C_2$ and $C_{3v}$ symmetries. The cases of the $C_{2v}$, $C_{4v}$ symmetries are worked out in appendix H below.

## 19.5 $C_2$ factorization

As the simplest example of implementing the above scheme consider the $C_2$ 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 $C_2$ symmetric. The $C_2$ symmetry cycles separate into two classes, the self-dual configurations $+-$, $++--$, $+++---$, $+--++-+-$, $\cdots$, with multiplicity $m_p = 1$, and the asymmetric configurations $+$, $-$, $++-$, $--+$, $\cdots$, with multiplicity $m_p = 2$. 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 19.4]

The symmetry reduced labeling $\rho_i \in \{0, 1\}$ is related to the standard $s_i \in \{+, -\}$ Ising spin labeling by

$$\begin{array}{llllll} \text{If} & s_i & = & s_{i-1} & \text{then} & \rho_i = 1 \\ \text{If} & s_i & \neq & s_{i-1} & \text{then} & \rho_i = 0 \end{array} \tag{19.18}$$

For example, $\overline{+} = \cdots + + + + \cdots$ maps into $\cdots 111 \cdots = \overline{1}$ (and so does $\overline{-}$), $\overline{-+} = \cdots - + - + \cdots$ maps into $\cdots 000 \cdots = \overline{0}$, $\overline{-++-} = \cdots - - + + - - + + \cdots$ maps into $\cdots 0101 \cdots = \overline{01}$, and so forth. A list of such reductions is given in table **??**.

Depending on the maximal symmetry group $\mathcal{H}_p$ that leaves an orbit $p$ invariant (see sects. 19.2 and 19.3 as well as sect. 19.1.1), the contributions to the dynamical zeta function factor as

$$\begin{array}{rrcl} & & & A_1 \quad A_2 \\ \mathcal{H}_p = \{e\}: & (1 - t_{\tilde{p}})^2 & = & (1 - t_{\tilde{p}})(1 - t_{\tilde{p}}) \\ \mathcal{H}_p = \{e, \sigma\}: & (1 - t_{\tilde{p}}^2) & = & (1 - t_{\tilde{p}})(1 + t_{\tilde{p}}) \ , \end{array} \tag{19.19}$$

For example:

$$\begin{array}{rrcl} \mathcal{H}_{++-} = \{e\}: & (1 - t_{++-})^2 & = & (1 - t_{001})(1 - t_{001}) \\ \mathcal{H}_{+-} = \{e, \sigma\}: & (1 - t_{+-}) & = & (1 - t_0)\ (1 + t_0), \quad t_{+-} = t_0^2 \end{array}$$

This yields two binary cycle expansions. The $A_1$ subspace dynamical zeta function is given by the standard binary expansion (18.7). The antisymmetric $A_2$ subspace dynamical zeta function $\zeta_{A_2}$ differs from $\zeta_{A_1}$ only by a minus sign for cycles with an odd number of 0's:

$$\begin{aligned} 1/\zeta_{A_2} &= (1 + t_0)(1 - t_1)(1 + t_{10})(1 - t_{100})(1 + t_{101})(1 + t_{1000}) \\ &\quad (1 - t_{1001})(1 + t_{1011})(1 - t_{10000})(1 + t_{10001}) \\ &\quad (1 + t_{10010})(1 - t_{10011})(1 - t_{10101})(1 + t_{10111}) \ldots \\ &= 1 + t_0 - t_1 + (t_{10} - t_1 t_0) - (t_{100} - t_{10} t_0) + (t_{101} - t_{10} t_1) \\ &\quad -(t_{1001} - t_1 t_{001} - t_{101} t_0 + t_{10} t_0 t_1) - \ldots\ldots \end{aligned} \tag{19.20}$$

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. 19.3.1) with group-theoretic factor $\mathbf{h}_p = (\mathbf{e} + \sigma)/2$, the boundary orbit does not contribute to the antisymmetric subspace

$$\begin{array}{rrcl} & & & A_1 \quad A_2 \\ \text{boundary:} & (1 - t_p) & = & (1 - t_{\tilde{p}})(1 - 0 t_{\tilde{p}}) \end{array} \tag{19.21}$$

This is the $1/\zeta$ part of the boundary orbit factorization of sect. 19.1.1.

## 19.6 $C_{3v}$ factorization: 3-disk game of pinball

The next example, the $C_{3v}$ symmetry, can be worked out by a glance at figure 11.2 (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 11.2 (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, $\overline{23} = \cdots 232323 \cdots$ maps into $\cdots 000 \cdots = \overline{0}$ (and so do $\overline{12}$ and $\overline{13}$), $\overline{123} = \cdots 12312 \cdots$ maps into $\cdots 111 \cdots = \overline{1}$ (and so does $\overline{132}$), and so forth. A list of such reductions for short cycles is given in table **??**.

$C_{3v}$ has two 1-dimensional irreducible representations, symmetric and anti-symmetric 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 (19.14) factorizes according to

$$\det(1-D(h)t) = (1-\chi_{A_1}(h)t)\,(1-\chi_{A_2}(h)t)\left(1-\chi_E(h)t+\chi_{A_2}(h)t^2\right)^2 \tag{19.22}$$

with the three factors contributing to the $C_{3v}$ irreducible representations $A_1$, $A_2$ and $E$, respectively, and the 3-disk dynamical zeta function factorizes into $\zeta = \zeta_{A_1}\zeta_{A_2}\zeta_E^2$. Substituting the $C_{3v}$ characters [10]

| $C_{3v}$ | $A_1$ | $A_2$ | $E$ |
|---|---|---|---|
| $e$ | 1 | 1 | 2 |
| $C, C^2$ | 1 | 1 | $-1$ |
| $\sigma_v$ | 1 | $-1$ | 0 |

into (19.22), we obtain for the three classes of possible orbit symmetries (indicated in the first column)

$$\begin{array}{rlcl} & \mathbf{h}_{\tilde{p}} & & A_1 \quad A_2 \quad E \\ e: & (1-t_{\tilde{p}})^6 & = & (1-t_{\tilde{p}})(1-t_{\tilde{p}})(1-2t_{\tilde{p}}+t_{\tilde{p}}^2)^2 \\ C, C^2: & (1-t_{\tilde{p}}^3)^2 & = & (1-t_{\tilde{p}})(1-t_{\tilde{p}})(1+\ t_{\tilde{p}}+t_{\tilde{p}}^2)^2 \\ \sigma_v: & (1-t_{\tilde{p}}^2)^3 & = & (1-t_{\tilde{p}})(1+t_{\tilde{p}})(1+0t_{\tilde{p}}-t_{\tilde{p}}^2)^2. \end{array} \tag{19.23}$$

where $\sigma_v$ stands for any one of the three reflections.

The Euler product (17.15) on each irreducible subspace follows from the factorization (19.23). On the symmetric $A_1$ subspace the $\zeta_{A_1}$ is given by the standard binary curvature expansion (18.7). The antisymmetric $A_2$ subspace $\zeta_{A_2}$ differs from $\zeta_{A_1}$ only by a minus sign for cycles with an odd number of 0's, and is given in (19.20). For the mixed-symmetry subspace $E$ the curvature expansion is given by

$$\begin{aligned}
1/\zeta_E &= (1+zt_1+z^2t_1^2)(1-z^2t_0^2)(1+z^3t_{100}+z^6t_{100}^2)(1-z^4t_{10}^2) \\
&\quad (1+z^4t_{1001}+z^8t_{1001}^2)(1+z^5t_{10000}+z^{10}t_{10000}^2) \\
&\quad (1+z^5t_{10101}+z^{10}t_{10101}^2)(1-z^5t_{10011})^2\ldots \\
&= 1+zt_1+z^2(t_1^2-t_0^2)+z^3(t_{001}-t_1t_0^2) \\
&\quad +z^4\left[t_{0011}+(t_{001}-t_1t_0^2)t_1-t_{01}^2\right] \\
&\quad +z^5\left[t_{00001}+t_{01011}-2t_{00111}+(t_{0011}-t_{01}^2)t_1+(t_1^2-t_0^2)t_{100}\right]
\end{aligned} \tag{19.24}$$

We have reinserted the powers of $z$ in order to group together cycles and pseudocycles 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 pseudocycles.

Referring back to the topological polynomial (13.31) obtained by setting $t_p = 1$, we see that its factorization is a consequence of the $C_{3v}$ factorization of the $\zeta$ function:

$$1/\zeta_{A_1} = 1-2z\ , \quad 1/\zeta_{A_2} = 1\ , \quad 1/\zeta_E = 1+z\ , \tag{19.25}$$

as obtained from (18.7), (19.20) and (19.24) for $t_p = 1$.

Their symmetry is $K = \{\mathbf{e}, \sigma\}$, so according to (19.11), they pick up the group-theoretic factor $\mathbf{h}_p = (\mathbf{e}+\sigma)/2$. If there is no sign change in $t_p$, then evaluation of $\det(1-\frac{\mathbf{e}+\sigma}{2}t_{\tilde{p}})$ yields

$$\text{boundary:}\quad (1-t_p)^3 = \overset{A_1}{(1-t_{\tilde{p}})}\overset{A_2}{(1-0t_{\tilde{p}})}\overset{E}{(1-t_{\tilde{p}})^2}\ , \quad t_p = t_{\tilde{p}}\ . \tag{19.26}$$

However, if the cycle weight changes sign under reflection, $t_{\sigma\tilde{p}} = -t_{\tilde{p}}$, the boundary orbit does not contribute to the subspace symmetric under reflection across the orbit;

$$\text{boundary:}\quad (1-t_p)^3 = \overset{A_1}{(1-0t_{\tilde{p}})}\overset{A_2}{(1-t_{\tilde{p}})}\overset{E}{(1-t_{\tilde{p}})^2}\ , \quad t_p = t_{\tilde{p}}\ . \tag{19.27}$$

## 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 length, 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 length 8, but reduction to the fundamental domain non-degenerate prime cycles reduces the number of the distinct cycles of length 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 length, 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 length 7 or less in full space, corresponding to 47 distinct periodic points. In the fundamental domain, we have 8 (distinct) periodic orbits up to length 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 (19.23) of the dynamical zeta function is *intrinsic* to the classical 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 [2]. In conclusion, the manifold advantages of the symmetry reduced dynamics should thus be obvious; full state space cycle expansions, such as those of exercise 18.8, are useful only for cross checking purposes.

## Commentary

**Remark 19.1** Symmetry reductions in periodic orbit theory. This chapter is based on long collaborative effort with B. Eckhardt, ref. [1]. The group-theoretic factorizations of dynamical zeta functions that we develop here were first introduced and applied in ref. [4]. They are closely related to the symmetrizations introduced by Gutzwiller [4] in the context of the semiclassical periodic orbit trace formulas, put into more general group-theoretic context by Robbins [2], whose exposition, together with Lauritzen's [3] treatment of the boundary orbits, has influenced the presentation given here. The symmetry reduced trace formula for a finite symmetry group $G = \{e, g_2, \ldots, g_{|G|}\}$ with $|G|$ group elements, where the integral over Haar measure is replaced by a finite group discrete sum $|G|^{-1} \sum_{g\in G} = 1$,

was derived in ref. [1]. A related group-theoretic decomposition in context of hyperbolic billiards was utilized in ref. [10], and for the Selberg's zeta function in ref. [11]. 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 irreducible representations of the corresponding Galois group.

**Remark 19.2** Computations. The techniques of this chapter have been applied to computations of the 3-disk classical and quantum spectra in refs. [7, 13], and to a "Zeeman effect" pinball and the $x^2y^2$ potentials in ref. [12]. 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 [10, 3, 13] for hyperbolic billiards [12] and Selberg zeta functions [18].

**Remark 19.3** Other symmetries. In addition to the symmetries exploited here, time reversal symmetry and a variety of other non-trivial discrete symmetries can induce further relations among orbits; we shall point out several of examples of cycle degeneracies under time reversal. We do not know whether such symmetries can be exploited for further improvements of cycle expansions.

## Exercises

19.1. **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 9.1. Compute the dynamical zeta function and the spectral determinant of the Perron-Frobenius operator for this map; check explicitly the factorization (19.2).

19.2. **2-$d$ asymmetric representation.** The above expressions can sometimes be simplified further using standard group-theoretical methods. For example, the $\frac{1}{2}\left((\mathrm{tr}\, M)^2 - \mathrm{tr}\, M^2\right)$ term in (19.16) is the trace of the antisymmetric part of the $M\times M$ Kronecker product. Show that if $\alpha$ is a 2-dimensional representation, this is the $A_2$ antisymmetric representation, and

$$\text{2-dim:} \quad \det(1-D_\alpha(h)t) = 1-\chi_\alpha(h)t+\chi_{A_2}(h)t^2. \tag{19.28}$$

19.3. **3-disk desymmetrization.**

a) 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).

b) Find the shortest cycle with no symmetries and factorize it as in a)

c) Find the shortest cycle that has the property that its time reversal is not described by the same symbolic dynamics.

d) Compute the dynamical zeta functions and the spectral determinants (symbolically) in the three representations; check the factorizations (19.15) and (19.17).

(Per Rosenqvist)

19.4. **$C_2$ factorizations: the Lorenz and Ising systems.** In the Lorenz system [1, 3] the labels + and − stand for the left or the right lobe of the attractor and the symmetry is a rotation by $\pi$ 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.

19.5. **Ising model.** The Ising model with two states $\epsilon_i = \{+,-\}$ per site, periodic boundary condition, and Hamil-

tonian

$$H(\epsilon) = -J \sum_i \delta_{\epsilon_i,\epsilon_{i+1}} ,$$

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.

19.6. **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

$$\det\left(1 - D(h)\frac{t_p}{\lambda_p^k}\right) ,$$

where $D(h)$ is the representation of $h$ in the regular representation of the group.

## References


[19.1] P. Cvitanović and B. Eckhardt, "Symmetry decomposition of chaotic dynamics," *Nonlinearity* **6**, 277 (1993).

[19.2] J.M. Robbins, "Semiclassical trace formulas in the presence of continuous symmetries," *Phys. Rev.* **A 40**, 2128 (1989).

[19.3] B. Lauritzen, Discrete symmetries and the periodic-orbit expansions, *Phys. Rev.* **A 43** 603, (1991).

[19.4] B. Eckhardt, G. Hose and E. Pollak, *Phys. Rev.* **A 39**, 3776 (1989).

[19.5] C. C. Martens, R. L. Waterland, and W. P. Reinhardt, *J. Chem. Phys.* **90**, 2328 (1989).

[19.6] S.G. Matanyan, G.K. Savvidy, and N.G. Ter-Arutyunyan-Savvidy, *Sov. Phys. JETP 53*, 421 (1981).

[19.7] A. Carnegie and I. C. Percival, *J. Phys.* **A 17**, 801 (1984).

[19.8] B. Eckhardt and D. Wintgen, *J. Phys.* **B 23**, 355 (1990).

[19.9] J.M. Robbins, S.C. Creagh and R.G. Littlejohn, *Phys. Rev.* **A39**, 2838 (1989); **A41**, 6052 (1990).

[19.10] M. Hamermesh, *Group Theory and its Application to Physical Problems* (Addison-Wesley, Reading, 1962).

[19.11] A. B. Venkov and P. G. Zograf, "Analogues of Artin's factorization formulas in the spectral theory of automorphic functions associated with induced representations of Fuchsian groups," *Math. USSR* **21**, 435 (1983).

[19.12] M.C. Gutzwiller, *J. Math. Phys.* **8**, 1979 (1967); **10**, 1004 (1969); **11**, 1791 (1970); **12**, 343 (1971).

[19.13] P. Scherer, *Quantenzustände eines klassisch chaotischen Billards*, Ph.D. thesis, Univ. Köln (Berichte des Forschungszentrums Jülich 2554, ISSN 0366-0885, Jülich, Nov. 1991).

# Chapter 20

# 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.

## 20.1 Escape rates

We start by verifying the claim (15.11) that for a nice hyperbolic flow the trace of the evolution operator grows exponentially with time. Consider again the game of pinball of figure 1.1. Designate by $\mathcal{M}$ a state space region that encloses the three disks, say the surface of the table $\times$ all pinball directions. The fraction of initial points whose trajectories start out within the state space region $\mathcal{M}$ and recur within that region at the time $t$ is given by

$$\hat{\Gamma}_{\mathcal{M}}(t) = \frac{1}{|\mathcal{M}|} \int \int_{\mathcal{M}} dxdy\, \delta\Big(y - f^t(x)\Big) \ . \tag{20.1}$$

This quantity is eminently 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 the 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 1.9. 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

$$\hat{\Gamma}_n = \frac{1}{|\mathcal{M}|} \sum_i^{(n)} |\mathcal{M}_i| \,, \tag{20.2}$$

where $i$ is a binary label of the $i$th strip, and $|\mathcal{M}_i|$ is the area of the $i$th strip. The phase space volume is preserved by the flow, so the strips of survivors are contracted along the stable eigendirections, and ejected along the unstable eigendirections. As a crude estimate of the number of survivors in the $i$th strip, assume that the spreading of a ray of trajectories per bounce is given by a factor $\Lambda$, the mean value of the expanding eigenvalue of the corresponding fundamental matrix of the flow, and replace $|\mathcal{M}_i|$ by the phase space strip width estimate $|\mathcal{M}_i|/|\mathcal{M}| \sim 1/\Lambda_i$. This estimate of a size of a neighborhood (given already on p.85) is right in spirit, but not without drawbacks. One problem is that in general the eigenvalues of a fundamental matrix for a finite segment of a trajectory have no invariant meaning; they depend on the choice of coordinates. However, we saw in chapter 16 that the sizes of neighborhoods are determined by stability eigenvalues of periodic points, and those are invariant under smooth coordinate transformations.

In the approximation $\hat{\Gamma}_n$ receives $2^n$ contributions of equal size

$$\hat{\Gamma}_1 \sim \frac{1}{\Lambda} + \frac{1}{\Lambda}, \cdots \qquad , \hat{\Gamma}_n \sim \frac{2^n}{\Lambda^n} = e^{-n(\lambda - h)} = e^{-n\gamma} \,, \tag{20.3}$$

up to pre-exponential factors. We see here the interplay of the two key ingredients of chaos first alluded to in sect. 1.3.1: the escape rate $\gamma$ equals local expansion rate (the Lyapunov exponent $\lambda = \ln \Lambda$), minus the rate of global reinjection back into the system (the topological entropy $h = \ln 2$).

As at each bounce one loses routinely the same fraction of trajectories, one expects the sum (20.2) to fall off exponentially with $n$. More precisely, by the hyperbolicity assumption of sect. 16.1.1 the expanding eigenvalue of the fundamental matrix of the flow is exponentially bounded from both above and below,

$$1 < |\Lambda_{min}| \leq |\Lambda(x)| \leq |\Lambda_{max}| \,, \tag{20.4}$$

and the area of each strip in (20.2) is bounded by $|\Lambda_{max}^{-n}| \leq |\mathcal{M}_i| \leq |\Lambda_{min}^{-n}|$. Replacing $|\mathcal{M}_i|$ in (20.2) by its over (under) estimates in terms of $|\Lambda_{max}|$, $|\Lambda_{min}|$ immediately leads to exponential bounds $(2/|\Lambda_{max}|)^n \leq \hat{\Gamma}_n \leq (2/|\Lambda_{min}|)^n$, i.e.,

$$\ln|\Lambda_{max}| - \ln 2 \geq \; -\frac{1}{n} \ln \hat{\Gamma}_n \; \geq \ln|\Lambda_{min}| - \ln 2 \,. \tag{20.5}$$

The argument based on (20.5) 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 $\gamma_n$ converge to the

limit $\gamma$, we first show that for hyperbolic systems the sum over survivor intervals (20.2) can be replaced by the sum over periodic orbit stabilities. By (20.4) the size of $\mathcal{M}_i$ strip can be bounded by the stability $\Lambda_i$ of $i$th periodic point:

$$C_1 \frac{1}{|\Lambda_i|} < \frac{|\mathcal{M}_i|}{|\mathcal{M}|} < C_2 \frac{1}{|\Lambda_i|}\,, \tag{20.6}$$

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 longer and longer cycles in a system of bounded hyperbolicity, the shrinking of the $i$th strip is better and better approximated by the derivatives evaluated on the periodic point within the strip. Hence the survival probability can be bounded close to the cycle point stability sum

$$\hat{C}_1\, \Gamma_n < \sum_i^{(n)} \frac{|\mathcal{M}_i|}{|\mathcal{M}|} < \hat{C}_2\, \Gamma_n \ , \tag{20.7}$$

where $\Gamma_n = \sum_i^{(n)} 1/|\Lambda_i|$ is the asymptotic trace sum (16.26). In this way we have established that for hyperbolic systems the survival probability sum (20.2) can be replaced by the periodic orbit sum (16.26).

[exercise 20.1]
[exercise 14.4]

We conclude that for hyperbolic, locally unstable flows the fraction (20.1) 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 $\gamma$ is the asymptotic *escape rate* defined by

$$\gamma = -\lim_{t\to\infty} \frac{1}{t} \ln \Gamma_{\mathcal{M}}(t)\,. \tag{20.8}$$

## 20.2 Natural measure in terms of periodic orbits

We now refine the reasoning of sect. 20.1. Consider the trace (16.7) in the asymptotic limit (16.25):

$$\mathrm{tr}\,\mathcal{L}^n = \int dx\, \delta(x - f^n(x))\, e^{\beta A^n(x)} \approx \sum_i^{(n)} \frac{e^{\beta A^n(x_i)}}{|\Lambda_i|}\,.$$

The factor $1/|\Lambda_i|$ was interpreted in (20.2) as the area of the $i$th phase space strip. Hence $\mathrm{tr}\,\mathcal{L}^n$ is a discretization of the *integral* $\int dx e^{\beta A^n(x)}$ approximated by a tessellation into strips centered on periodic points $x_i$, figure 1.11, with the volume

of the $i$th neighborhood given by estimate $|\mathcal{M}_i| \sim 1/|\Lambda_i|$, and $e^{\beta A^n(x)}$ estimated by $e^{\beta A^n(x_i)}$, its value at the $i$th periodic point. If the symbolic dynamics is a complete, any rectangle $[s_{-m} \cdots s_0.s_1 s_2 \cdots s_n]$ of sect. 11.4.1 always contains the cycle point $\overline{s_{-m} \cdots s_0 s_1 s_2 \cdots s_n}$; hence even though the periodic points are of measure zero (just like rationals in the unit interval), they are dense on the non–wandering set. Equipped with a measure for the associated rectangle, periodic orbits suffice to cover the entire non–wandering set. The average of $e^{\beta A^n}$ evaluated on the non–wandering set is therefore given by the trace, properly normalized so $\langle 1 \rangle = 1$:

$$\left\langle e^{\beta A^n} \right\rangle_n \approx \frac{\sum_i^{(n)} e^{\beta A^n(x_i)}/|\Lambda_i|}{\sum_i^{(n)} 1/|\Lambda_i|} = \sum_i^{(n)} \mu_i \, e^{\beta A^n(x_i)} \,. \tag{20.9}$$

Here $\mu_i$ is the *normalized natural measure*

$$\sum_i^{(n)} \mu_i = 1 \,, \qquad \mu_i = e^{n\gamma}/|\Lambda_i| \,, \tag{20.10}$$

correct both for the closed systems as well as the open systems of sect.15.1.3.

Unlike brute numerical slicing of the integration space into an arbitrary lattice (for a critique, see sect. 14.3), the periodic orbit theory is smart, as it automatically partitions integrals by the intrinsic topology of the flow, and assigns to each tile the invariant natural measure $\mu_i$.

### 20.2.1 Unstable periodic orbits are dense

(L. Rondoni and P. Cvitanović)

Our goal in sect. 15.1 was to evaluate the space and time averaged expectation value (15.9). An average over all periodic orbits can accomplish the job only if the periodic orbits fully explore the asymptotically accessible state space.

Why should the 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 pass infinitely many times arbitrarily close to any point on the set, including the initial point of the trajectory itself. The argument is more or less the following. Take a partition of $\mathcal{M}$ in arbitrarily small regions, and consider particles that start out in region $\mathcal{M}_i$, and return to it in $n$ steps after some peregrination in state space. In particular, a particle might return a little to the left of its original position, while a close 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 larger and larger period 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.

## 20.3 Flow conservation sum rules

If the dynamical system is bounded, all trajectories remain confined for all times, escape rate (20.8) vanishes $\gamma = -s_0 = 0$, and the leading eigenvalue of the Perron-Frobenius operator (14.10) is simply $\exp(-t\gamma) = 1$. Conservation of material flow thus implies that for bound flows cycle expansions of dynamical zeta functions and spectral determinants satisfy exact *flow conservation* sum rules:

$$\begin{aligned} 1/\zeta(0,0) &= 1 + \sum_{\pi}{}' \frac{(-1)^k}{|\Lambda_{p_1}\cdots\Lambda_{p_k}|} = 0 \\ F(0,0) &= 1 - \sum_{n=1}^{\infty} c_n(0,0) = 0 \end{aligned} \tag{20.11}$$

obtained by setting $s = 0$ in (18.15), (18.16) cycle weights $t_p = e^{-sT_p}/|\Lambda_p| \to 1/|\Lambda_p|$ . 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 $\Lambda_{p,1}$, $\Lambda_{p,2}$, $\cdots$, $\Lambda_{p,d}$, and their significance is purely geometric: they are a measure of how well periodic orbits tessellate the state space. Conservation of material flow provides the first and very useful test of the quality of finite cycle length truncations, and is something that you should always check first when constructing a cycle expansion for a bounded flow.

The trace formula version of the flow conservation flow sum rule comes in two varieties, one for the maps, and another for the flows. By flow conservation the leading eigenvalue is $s_0 = 0$, and for maps (18.14) yields

$$\mathrm{tr}\,\mathcal{L}^n = \sum_{i\in \mathrm{Fix} f^n} \frac{1}{|\det\left(\mathbf{1} - M^n(x_i)\right)|} = 1 + e^{s_1 n} + \dots \ . \tag{20.12}$$

For flows one can apply this rule by grouping together cycles from $t = T$ to $t = T + \Delta T$

$$\begin{aligned} \frac{1}{\Delta T} \sum_{p,r}^{T\le rT_p\le T+\Delta T} \frac{T_p}{\left|\det\left(\mathbf{1} - M_p^r\right)\right|} &= \frac{1}{\Delta T}\int_T^{T+\Delta T} dt\left(1 + e^{s_1 t} + \dots\right) \\ &= 1 + \frac{1}{\Delta T}\sum_{\alpha=1}^{\infty} \frac{e^{s_\alpha T}}{s_\alpha}\left(e^{s_\alpha \Delta T} - 1\right) \approx 1 + e^{s_1 T} + \cdots . \end{aligned} \tag{20.13}$$

As is usual for the the fixed level trace sums, the convergence of (20.12) is controled by the gap between the leading and the next-to-leading eigenvalues of the evolution operator.

## 20.4 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

$$C_{AB}(t; x_0) \;=\; \lim_{T\to\infty} \frac{1}{T} \int_0^T d\tau A(x(\tau + t))B(x(\tau)) \,, \qquad x_0 = x(0) \,. \quad (20.14)$$

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

$$C_{AB}(t) \;=\; \int_{\mathcal{M}} dx_0\, \rho_0(x_0) A(f^t(x_0))B(x_0) \,. \quad (20.15)$$

For a chaotic system we expect that time evolution will loose 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

$$\hat{C}_{AB} := C_{AB} - \langle A\rangle \langle B\rangle \quad (20.16)$$

for any pair of observables coincides with the definition of *mixing*, a fundamental property in ergodic theory. We now assume $\langle B\rangle = 0$ (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 (20.15) 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

$$\tilde{C}_{AB}(t) \;=\; \int_{\mathcal{M}} dx \int_{\mathcal{M}} dy\, A(y)\mathcal{L}^t(y, x)B(x)\rho_0(x)$$

We recall that in sect. 14.1 we showed that $\rho(x)$ is the eigenvector of $\mathcal{L}$ corresponding to probability conservation

$$\int_{\mathcal{M}} dy\, \mathcal{L}^t(x, y)\rho(y) \;= \rho(x) \,.$$

Now, we can expand the $x$ dependent part 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 $\mathcal{L}$ then can be written as

$$\tilde{C}_{AB}(t) \;=\; \sum_{\alpha\neq 0} e^{-s_\alpha t} c_\alpha \int_{\mathcal{M}} dy\, A(y)\rho_\alpha(y). \tag{20.17}$$

[exercise 20.2]

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 (20.17) implies *exponential* decay of correlations

$$\tilde{C}_{AB}(t) \sim e^{-\nu t}\,.$$

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), while 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 linked to transport exponents.

## 20.5 Trace formulas *vs.* level sums

Trace formulas (16.10) and (16.23) diverge precisely where one would like to use them, at $s$ equal to eigenvalues $s_\alpha$. Instead, one can proceed as follows; according to (16.27) the "level" sums (all symbol strings of length $n$) are asymptotically going like $e^{s_0 n}$

$$\sum_{i\in \mathrm{Fix} f^n} \frac{e^{\beta A^n(x_i)}}{|\Lambda_i|} \to e^{s_0 n}\,,$$

so an $n$th order estimate $s_{(n)}$ of the leading eigenvalue is given by

$$1 = \sum_{i\in \mathrm{Fix} f^n} \frac{e^{\beta A^n(x_i)} e^{-s_{(n)} n}}{|\Lambda_i|} \tag{20.18}$$

which generates a "normalized measure." The difficulty with estimating this $n \to \infty$ limit is 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 $s_{(n)}$ is not unique, because the sums $\Gamma_n$ depend on how we define the escape region, and because in general the areas $\mathcal{M}_i$ in the sum (20.2) should be weighted by the density of initial conditions $x_0$. For example, an overall measuring unit rescaling $\mathcal{M}_i \to \alpha\mathcal{M}_i$ introduces $1/n$ corrections in $s_{(n)}$ defined by the log of the sum (20.8): $s_{(n)} \to s_{(n)} - \ln\alpha/n$. This can be partially fixed by defining a level average

$$\left\langle e^{\beta A(s)} \right\rangle_{(n)} := \sum_{i\in \mathrm{Fix}\, f^n} \frac{e^{\beta A^n(x_i)} e^{sn}}{|\Lambda_i|} \tag{20.19}$$

and requiring that the ratios of successive levels satisfy

$$1 = \frac{\left\langle e^{\beta A(s_{(n)})} \right\rangle_{(n+1)}}{\left\langle e^{\beta A(s_{(n)})} \right\rangle_{(n)}} .$$

This avoids the worst problem with the formula (20.18), the inevitable $1/n$ 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: the dynamical zeta functions and spectral determinants are already invariant not only under linear rescalings, but under *all* smooth nonlinear conjugacies $x \to h(x)$, and require no $n \to \infty$ extrapolations to asymptotic times. Comparing with the cycle expansions (18.7) we see what the difference is; while in the level sum approach we keep increasing exponentially the number of terms with no reference to the fact that most are already known from shorter estimates, in the cycle expansions short terms dominate, longer ones enter only as exponentially small corrections.

The beauty of the trace formulas is that they are coordinatization independent: both $\left|\det\left(\mathbf{1} - M_p\right)\right| = |\det\left(\mathbf{1} - M^{T_p}(x)\right)|$ and $e^{\beta A_p} = e^{\beta A^{T_p}(x)}$ contribution to the cycle weight $t_p$ are independent of the starting periodic point point $x$. For the fundamental matrix $M_p$ this follows from the chain rule for derivatives, and for $e^{\beta A_p}$ from the fact that the integral over $e^{\beta A^t(x)}$ is evaluated along a closed loop. In addition, $\left|\det\left(\mathbf{1} - M_p\right)\right|$ is invariant under smooth coordinate transformations.

## Résumé

We conclude this chapter by a general comment on the relation of the finite trace sums such as (20.2) to the spectral determinants and dynamical zeta functions. One might be tempted to believe that given a deterministic rule, a sum like (20.2) could be evaluated to any desired precision. For short finite times this is indeed true: every region $\mathcal{M}_i$ in (20.2) can be accurately delineated, and there is no need

for fancy theory. However, if the dynamics is unstable, local variations in initial conditions grow exponentially and in finite time attain the size of the system. The difficulty with estimating the $n \to \infty$ limit from (20.2) 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 $\gamma_n$ is not unique, because the sums $\hat{\Gamma}_n$ depend on how we define the escape region, and because in general the areas $|\mathcal{M}_i|$ in the sum (20.2) should be weighted by the density of initial $x_0$.

In contrast, the dynamical zeta functions and spectral determinants are invariant under *all* smooth nonlinear conjugacies $x \to h(x)$, not only linear rescalings, and require no $n \to \infty$ extrapolations.

## Commentary

**Remark 20.1** Nonhyperbolic measures. $\mu_i = 1/|\Lambda_i|$ is the natural measure only for the strictly hyperbolic systems. For non-hyperbolic systems, the measure might develop cusps. For example, for Ulam type maps (unimodal maps with quadratic critical point mapped onto the "left" unstable fixed point $x_0$, discussed in more detail in chapter 23), the measure develops a square-root singularity on the $\overline{0}$ cycle:

$$\mu_0 = \frac{1}{|\Lambda_0|^{1/2}} \ . \tag{20.20}$$

The thermodynamics averages are still expected to converge in the "hyperbolic" phase where the positive entropy of unstable orbits dominates over the marginal orbits, but they fail in the "non-hyperbolic" phase. The general case remains unclear [19, 2, 3, 5].

**Remark 20.2** Trace formula periodic orbit averaging. The cycle averaging formulas are not the first thing that one would intuitively write down; the approximate trace formulas are more accessibly heuristically. The trace formula averaging (20.13) seems to have be discussed for the first time by Hannay and Ozorio de Almeida [8, 11]. Another novelty of the cycle averaging formulas and one of their main virtues, in contrast to the explicit analytic results such as those of ref. [4], is that their evaluation *does not* require any explicit construction of the (coordinate dependent) eigenfunctions of the Perron-Frobenius operator (i.e., the natural measure $\rho_0$).

**Remark 20.3** Role of noise in dynamical systems. In any physical application the dynamics is always accompanied by additional external noise. The noise can be characterized by its strength $\sigma$ and distribution. Lyapunov exponents, correlation decay and dynamo rate can be defined in this case the same way as in the deterministic case. You might fear that noise completely destroys the results derived here. However, one can show that the 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, with deterministically non-communicating parts of the state space now weakly

connected due to the noise, making periodic orbit theory applicable to non-ergodic systems. For small amplitude noise one can expand

$$\overline{a} = \overline{a}_0 + \overline{a_1}\sigma^2 + \overline{a_2}\sigma^4 + \dots ,$$

around the deterministic averages $a_0$. The expansion coefficients $\overline{a_1}, \overline{a_2}, \dots$ can also be expressed via periodic orbit formulas. The calculation of these coefficients is one of the challenges facing periodic orbit theory, discussed in refs. [5, 6, 7].

## Exercises

20.1. **Escape rate of the logistic map.**

(a) Calculate the fraction of trajectories remaining trapped in the interval [0, 1] for the logistic map

$$f(x) = A(1 - (2x - 1)^2), \tag{20.21}$$

and determine the $A$ dependence of the escape rate $\gamma(A)$ numerically.

(b) Work out a numerical method for calculating the lengths of intervals of trajectories remaining stuck for $n$ iterations of the map.

(c) What is your expectation about the $A$ dependence near the critical value $A_c = 1$?

20.2. **Four scale map decay.** Compute the second largest eigenvalue of the Perron-Frobenius operator for the four scale map

$$f(x) = \begin{cases} a_1 x & \text{if } 0 < x < b/a_1 \\ (1-b)((x - b/a_1)/(b - b/a_1)) + b & \text{if } b/a_1 < x < b \\ a_2(x-b) & \text{if } b < x < b + b/a_2, \\ (1-b)((x - b - b/a_2)/(1 - b - b/a_2)) + b & \text{if } b + b/a_2 < x < 1 \end{cases} \tag{20.22}$$

20.3. **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 some or all of the following maps:

(a) the piecewise-linear flow conserving map, the skew tent map

$$f(x) = \begin{cases} ax & \text{if } 0 \le x \le a^{-1}, \\ \frac{a}{a-1}(1-x) & \text{if } a^{-1} \le x \le 1. \end{cases}$$

(b) the Ulam map $f(x) = 4x(1-x)$

(c) the skew Ulam map

$$f(x) = \Lambda_0 x(1-x)(1-bx) , \quad 1/\Lambda_0 = x_c(1-x_c)(1-bx_c) . \tag{20.23}$$

In our numerical work we fix (arbitrarily, the value chosen in ref. [3]) $b = 0.6$, so

$$f(x) = 0.1218\, x(1-x)(1-0.6\, x)$$

with a peak at 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 18.2).

(e) for the 2-branch flow conserving map

$$f_0(x) = \frac{h - p + \sqrt{(h-p)^2 + 4hx}}{2h}, \qquad x \in [0$$

$$f_1(x) = \frac{h + p - 1 + \sqrt{(h+p-1)^2 + 4h(x-p}}{2h}$$

This is a nonlinear perturbation of ($h = 0$) Bernoulli map (21.6); the first 15 eigenvalues of the Perron-Frobenius operator are listed in ref. [1] 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 cross check your cycle expansions results for cases (c) and (d) with Lyapunov exponent computed by direct numerical averaging along trajectories of randomly chosen initial points:

(f) trajectory-trajectory separation (15.27) (hint: rescale $\delta x$ every so often, to avoid numerical overflows),

(g) iterated stability (15.32).

How good is the numerical accuracy compared with the periodic orbit theory predictions? oo

## References

[20.1] F. Christiansen, G. Paladin and H.H. Rugh, *Phys. Rev. Lett.* **65**, 2087 (1990).

[20.2] A. Politi, R. Badii and P. Grassberger, *J. Phys. A* **15**, L763 (1988); P. Grassberger, R. Badii and A. Politi, "Scaling laws for invariant measures on hyperbolic and nonhyperbolic attractors," *J. Stat. Phys.* **51**, 135 (1988).

[20.3] E. Ott, C. Grebogi and J.A. Yorke, *Phys. Lett. A* **135**, 343 (1989).

[20.4] C. Grebogi, E. Ott and J.A. Yorke, *Phys. Rev.* **A 36**, 3522 (1987).

[20.5] C. Grebogi, E. Ott and J. Yorke, *Physica* **D 7**, 181 (1983).

[20.6] C. Grebogi, E. Ott and J.A. Yorke, *Phys. Rev.* **A36**, 3522 (1987).

[20.7] C. Grebogi, E. Ott and J. Yorke, "Unstable periodic orbits and the dimension of multifractal chaotic attractors," *Phys. Rev. A* **37**, 1711 (1988).

[20.8] J. H. Hannay and A. M. Ozorio de Almeida, *J. Phys.* **A 17**, 3429, (1984).

# Chapter 21

# 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 chapters 16 and 6 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 21.1 we compute explicitly the eigenspectrum for the three analytically tractable piecewise linear examples. In sect. 21.3 we review the basic facts of the classical Fredholm theory of integral equations. The program is sketched in sect. 21.4, motivated by an explicit study of eigenspectrum of the Bernoulli shift map, and in sect. 21.5 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 21.1]

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.

fast track:
chapter 12, p. 195

## 21.1 Linear maps: exact spectra

We start gently; in example 21.1 we work out the *exact* eigenvalues and eigenfunctions of the Perron-Frobenius operator for the simplest example of unstable, expanding dynamics, a linear 1-$d$ map with one unstable fixed point. . Ref. [6] shows that this can be carried over to $d$-dimensions. Not only that, but in example 21.5 we compute the exact spectrum for the simplest example of a dynamical system with an *infinity* of unstable periodic orbits, the Bernoulli shift.

**Example 21.1** ***The simplest eigenspectrum - a single fixed point:*** *In order to get some feeling for the determinants defined so formally in sect. 17.2, let us work out a trivial example: a repeller with only one expanding linear branch*

$$f(x) = \Lambda x\,, \qquad |\Lambda| > 1\,,$$

*and only one fixed point* $x^* = 0$. *The action of the Perron-Frobenius operator (14.10) is*

$$\mathcal{L}\phi(y) = \int dx\,\delta(y - \Lambda x)\,\phi(x) = \frac{1}{|\Lambda|}\phi(y/\Lambda)\,. \tag{21.1}$$

*From this one immediately gets that the monomials* $y^k$ *are eigenfunctions:*

$$\mathcal{L}y^k = \frac{1}{|\Lambda|\Lambda^k}y^k\,, \quad k = 0, 1, 2, \ldots \tag{21.2}$$

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 $1/|\Lambda|\Lambda^k$ is getting exponentially smaller because densities that vary more rapidly decay more rapidly under the expanding action of the map.

**Example 21.2** ***The trace formula for a single fixed point:*** *The eigenvalues* $\Lambda^{-k-1}$ *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 (16.7). A similar result follows for powers of $\mathcal{L}$, yielding the single-fixed point version of the trace formula for maps (16.10):*

$$\sum_{k=0}^{\infty} \frac{ze^{s_k}}{1-ze^{s_k}} = \sum_{r=1}^{\infty} \frac{z^r}{|1-\Lambda^r|}, \qquad e^{s_k} = \frac{1}{|\Lambda|\Lambda^k}. \tag{21.3}$$

The left hand side of (21.3) is a meromorphic function, with the leading zero at $z = |\Lambda|$. So what?

**Example 21.3** ***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*

$$h_N(z) = \sum_{k=0}^{N} \sigma_k z^k = \frac{a}{b} + \frac{z(a-b)(1-z^N/b^N)}{b^2(1-z/b)}.$$

*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.*

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 21.4** ***The spectral determinant for a single fixed point:*** *The spectral determinant (17.3) follows from the trace formulas of example 21.2:*

$$\det(1-z\mathcal{L}) = \prod_{k=0}^{\infty}\left(1-\frac{z}{|\Lambda|\Lambda^k}\right) = \sum_{n=0}^{\infty}(-t)^n Q_n, \quad t = \frac{z}{|\Lambda|}, \tag{21.4}$$

*where the cummulants $Q_n$ are given explicitly by the* Euler formula

[exercise 21.3]

$$Q_n = \frac{1}{1-\Lambda^{-1}}\,\frac{\Lambda^{-1}}{1-\Lambda^{-2}}\cdots\frac{\Lambda^{-n+1}}{1-\Lambda^{-n}}. \tag{21.5}$$

The main lesson to glean from this simple example is that the cummulants $Q_n$ decay asymptotically *faster* than exponentially, as $\Lambda^{-n(n-1)/2}$. For example, if we approximate series such as (21.4) 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 length? We now turn to the simplest example of a dynamical system with an infinity of unstable periodic orbits.

**Example 21.5** ***Bernoulli shift:*** *Consider next the Bernoulli shift map*

$$x \mapsto 2x \ \ (mod\ 1)\,, \qquad x \in [0,1]\,. \tag{21.6}$$

*The associated Perron-Frobenius operator (14.9) assambles $\rho(y)$ from its two preimages*

$$\mathcal{L}\rho(y) = \frac{1}{2}\rho\left(\frac{y}{2}\right) + \frac{1}{2}\rho\left(\frac{y+1}{2}\right)\,. \tag{21.7}$$

*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 generating function*

$$\mathcal{G}(x,t) \;=\; \frac{te^{xt}}{e^t-1} \;=\; \sum_{k=0}^{\infty} B_k(x)\frac{t^k}{k!}\,, \quad B_0(x)=1\,,\ B_1(x) = x - \frac{1}{2}\,,\ldots$$

*The Perron-Frobenius operator (21.7) acts on the 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 length $n$. Without further ado we substitute everything back and obtain the determinant,*

$$\det(1-z\mathcal{L}) = \exp\left(-\sum_{n=1}\frac{z^n}{n}\,\frac{2^n}{2^n-1}\right) = \prod_{k=0}\left(1-\frac{z}{2^k}\right)\,, \tag{21.8}$$

*verifying that the Bernoulli polynomials are eigenfunctions with eigenvalues $1, 1/2, \ldots, 1/2^n, \ldots$.*

The Bernoulli map spectrum looks reminiscent of the single fixed-point spectrum (21.2), 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 (1.6) 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 (17.23), for bound systems the local expansion rate (here $\ln|\Lambda| = \ln 2$) [section 17.4]

is balanced by the entropy (here $\ln 2$, the log of the number of preimages $F_s$), yielding zero escape rate.

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-$d$ hyperbolic map next.

**Example 21.6** ***The simplest of 2-$d$ 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 (14.10) acts on the 2-$d$ density functions as*

$$\mathcal{L}\rho(x_1, x_2) = \frac{1}{|\Lambda_s \Lambda_u|}\,\rho(x_1/\Lambda_s, x_2/\Lambda_u) \tag{21.9}$$

*What are good eigenfunctions? Cribbing the 1-$d$ eigenfunctions for the stable, contracting $x_1$ direction from example 21.1 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*

$$\varphi_{k_1 k_2}(x_1, x_2) = x_2^{k_2}/x_1^{k_1+1}\,, \qquad k_1,\, k_2 = 0, 1, 2, \ldots\,, \tag{21.10}$$

*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*

$$\mathcal{L}\varphi_{k_1 k_2}(x_1, x_2) \;=\; \frac{\sigma}{|\Lambda_u|}\,\frac{\Lambda_s^{k_1}}{\Lambda_u^{k_2}}\,\varphi_{k_1 k_2}(x_1, x_2)\,, \qquad \sigma = \Lambda_s/|\Lambda_s|$$

*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 (17.9)) 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 $f(x_1...., x_d) = (\Lambda_1 x_1, \Lambda_2 x_2, \ldots, \Lambda_d x_d)$.*

So far we have checked the trace and spectral determinant formulas derived heuristically in chapters 16 and 17, but only for the case of 1- and 2-$d$ linear maps. But for infinite-dimensional vector spaces this game is fraught with dangers, and we have already been mislead by piecewise linear examples into spectral confusions: contrast the spectra of example 14.1 and example 15.2 with the spectrum computed in example 16.1.

We show next that the above results do carry over to a sizable class of piecewise analytic expanding maps.

## 21.2 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. 14.3. The problem with such state space discretization approaches that they sometimes yield plainly wrong spectra (compare example 15.2 with the result of example 16.1), 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

$$\phi_n(y) = \sum_{k=0}^{\infty} \phi_n^{(k)}(0)\frac{y^k}{k!}\,, \qquad \phi_{n+1}(y)_k = \sum_{\ell=0}^{\infty} \phi_{n+1}^{(\ell)}(0)\frac{y^\ell}{\ell!}\,,$$

$$\phi_{n+1}^{(\ell)}(0) = \int dx\ \delta^{(\ell)}(y - f(x))\phi_n(x)\Big|_{y=0}\,, \qquad x = f^{-1}(0)\,,$$

and substitute the two Taylor series into (14.6):

$$\phi_{n+1}(y) = (\mathcal{L}\phi_n)(y) = \int_{\mathcal{M}} dx\, \delta(y - f(x))\, \phi_n(x)\,.$$

The matrix elements follow by evaluating the integral

$$\mathbf{L}_{\ell k} = \frac{\partial^\ell}{\partial y^\ell} \int dx\, \mathcal{L}(y,x)\frac{x^k}{k!}\Big|_{y=0}\,. \tag{21.11}$$

we obtain a matrix representation of the evolution operator

$$\int dx\, \mathcal{L}(y,x)\frac{x^k}{k!} = \sum_{k'} \frac{y^{k'}}{k'!}\mathbf{L}_{k'k}\,, \qquad k, k' = 0, 1, 2, \ldots$$

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. 14.2. This yields a representation of the evolution operator centered on the

fixed point, evaluated recursively in terms of derivatives of the map $f$:

$$
\begin{aligned}
(\mathbf{L})_{\ell k} &= \int dx\, \delta^{(\ell)}(x - f(x)) \frac{x^k}{k!}\bigg|_{x=f(x)} \\
&= \frac{1}{|f'|}\left(\frac{d}{dx}\frac{1}{f'(x)}\right)^{\ell} \frac{x^k}{k!}\bigg|_{x=f(x)} .
\end{aligned}
\tag{21.12}
$$

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

$$
\mathbf{L}_{kk} = \frac{1}{|\Lambda|\Lambda^k}\,, \quad \mathbf{L}_{k+1,k} = -\frac{(k+2)!\, f''}{2k!|\Lambda|\Lambda^{k+2}}\,, \quad \cdots .
$$

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 (21.3) takes now a matrix form

$$
\mathrm{tr}\,\frac{z\mathcal{L}}{1 - z\mathcal{L}} = \mathrm{tr}\,\frac{\mathbf{L}}{1 - z\mathbf{L}}\,. \tag{21.13}
$$

In order to illustrate how this works, we work out a few examples.

In example 21.7 we show that these results carry over to any analytic single-branch 1-$d$ repeller. Further examples motivate the steps that lead to a proof that spectral determinants for general analytic 1-dimensional expanding maps, and - in sect. 21.5, for 2-dimensional hyperbolic mappings - are also entire functions.

**Example 21.7** ***Perron-Frobenius operator in a matrix representation:*** *As in example 21.1, 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 $x = w^*$,*

$$
\mathcal{L}\phi(y) = \int dx\, \delta(y - f(x))\, \phi(x) = \sigma\, F'(y)\, \phi(F(y))\,.
$$

*Assume that $F$ is a contraction of the unit disk in the complex plane, i.e.,*

$$
|F(z)| < \theta < 1 \quad \textit{and} \quad |F'(z)| < C < \infty \quad \textit{for} \quad |z| < 1\,, \tag{21.14}
$$

*and expand $\phi$ in a polynomial basis with the Cauchy integral formula*

$$
\phi(z) = \sum_{n=0}^{\infty} z^n \phi_n = \oint \frac{dw}{2\pi i}\, \frac{\phi(w)}{w - z}\,, \quad \phi_n = \oint \frac{dw}{2\pi i}\, \frac{\phi(w)}{w^{n+1}}
$$

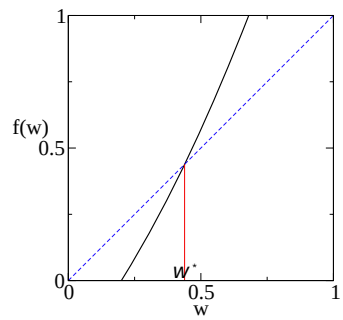



**Figure 21.1:** A nonlinear one-branch repeller with a single fixed point $w^*$.

*Combining this with (21.22), we see that in this basis Perron-Frobenius operator $\mathcal{L}$ is represented by the matrix*

$$\mathcal{L}\phi(w) = \sum_{m,n} w^m L_{mn}\phi_n \,, \quad L_{mn} = \oint \frac{dw}{2\pi i} \, \frac{\sigma\, F'(w)(F(w))^n}{w^{m+1}} \,. \tag{21.15}$$

*Taking the trace and summing we get:*

$$\mathrm{tr}\ \mathcal{L} = \sum_{n\geq 0} L_{nn} = \oint \frac{dw}{2\pi i} \, \frac{\sigma\, F'(w)}{w - F(w)} \,.$$

*This integral has but one simple pole at the unique fixed point $w^* = F(w^*) = f(w^*)$. Hence*

[exercise 21.6]

$$\mathrm{tr}\ \mathcal{L} = \frac{\sigma\, F'(w^*)}{1 - F'(w^*)} = \frac{1}{|f'(w^*) - 1|} \,.$$

This super-exponential decay of cummulants $Q_k$ ensures that for a repeller consisting of a single repelling point the spectral determinant (21.4) 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 classical 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 expensions approach.

Now, if the spectral determinant is entire, formulas such as (17.25) 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 (21.4) - super-exponentially to the exact values, and so the cycle expansions to be discussed in chapter 18 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. 21.4, p. 357

## 21.3 Classical 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 classical Fredholm integral operator

$$\mathcal{K}\varphi(x) \;=\; \int_{\mathcal{M}} dy\, \mathcal{K}(x,y)\varphi(y)\,, \tag{21.16}$$

and one is tempted to resort too classical 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 21.5.

The general form of Fredholm integral equations of the second kind is

$$\varphi(x) \;=\; \int_{\mathcal{M}} dy\, \mathcal{K}(x,y)\varphi(y) \;+\; \xi(x) \tag{21.17}$$

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 (21.16), 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 $\delta$, so we immediately realize (figure 21.4) that much work is needed to bring Perron-Frobenius operators into this picture.

We rewrite (21.17) in the form

$$\mathcal{T}\varphi = \xi\,, \qquad \mathcal{T} = \mathbb{1} - \mathcal{K}\,. \tag{21.18}$$

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 $\lambda$ 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 $\lambda$ is sufficiently small we may look for a perturbative expression for such an inverse, as a geometric series

$$(\mathbb{1} - \lambda\mathcal{K})^{-1} = \mathbb{1} + \lambda\mathcal{K} + \lambda^2\mathcal{K}^2 + \cdots = \mathbb{1} + \lambda\mathcal{W}\,, \tag{21.19}$$

where $\mathcal{K}^n$ is a compact integral operator with kernel

$$\mathcal{K}^n(x,y) = \int_{\mathcal{M}^{n-1}} dz_1 \ldots dz_{n-1}\, \mathcal{K}(x,z_1)\cdots\mathcal{K}(z_{n-1},y)\,,$$

and $\mathcal{W}$ is also compact, as it is given by the convergent sum of compact operators. The problem with (21.19) is that the series has a finite radius of convergence, while apart from a denumerable set of $\lambda$'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 $\lambda$

$$\mathcal{W}(x,y) = \frac{\mathcal{D}(x,y;\lambda)}{D(\lambda)}\,.$$

If we introduce the notation

$$\mathcal{K}\begin{pmatrix} x_1 \ldots x_n \\ y_1 \ldots y_n \end{pmatrix} = \begin{vmatrix} \mathcal{K}(x_1,y_1) & \ldots & \mathcal{K}(x_1,y_n) \\ \ldots & \ldots & \ldots \\ \mathcal{K}(x_n,y_1) & \ldots & \mathcal{K}(x_n,y_n) \end{vmatrix}$$

we may write the explicit expressions

$$\begin{aligned} D(\lambda) &= 1 + \sum_{n=1}^{\infty} (-1)^n \frac{\lambda^n}{n!} \int_{\mathcal{M}^n} dz_1 \ldots dz_n\, \mathcal{K}\begin{pmatrix} z_1 \ldots z_n \\ z_1 \ldots z_n \end{pmatrix} \\ &= \exp\left(-\sum_{m=1}^{\infty} \frac{\lambda^m}{m} \mathrm{tr}\,\mathcal{K}^m\right) \qquad (21.20) \\ \mathcal{D}(x,y;\lambda) &= \mathcal{K}\begin{pmatrix} x \\ y \end{pmatrix} + \sum_{n=1}^{\infty} \frac{(-\lambda)^n}{n!} \int_{\mathcal{M}^n} dz_1 \ldots dz_n\, \mathcal{K}\begin{pmatrix} x & z_1 & \ldots & z_n \\ y & z_1 & \ldots & z_n \end{pmatrix} \end{aligned}$$

The quantity $D(\lambda)$ is known as the Fredholm determinant (see (17.24)):it is an entire analytic function of $\lambda$, 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.

## 21.4 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,

$$\phi_\theta(y) = \sum_{k\neq 0} \exp(2\pi iky)\frac{1}{|k|^\theta} \tag{21.21}$$

with complex eigenvalue $2^{-\theta}$, parameterized by complex $\theta$ with Re $\theta > 0$. By varying $\theta$ one realizes that such eigenvalues fill out the entire unit disk. Such *essential spectrum*, the case $k = 0$ of figure 21.4, 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 $\mathbf{\Omega}$, the spectral radius is the smallest positive number $\rho_{spec}$ such that the spectrum is inside the disk of radius $\rho_{spec}$, while the essential spectral radius is the smallest positive number $\rho_{ess}$ such that outside the disk of radius $\rho_{ess}$ the spectrum consists only of isolated eigenvalues of finite multiplicity (see figure 21.4). [exercise 21.5]

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. 21.1. 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 $0 < \mathrm{Re}\,\theta < k + \alpha$ most $\phi_\theta$ will cease to be eigenfunctions in the space $\mathbb{C}^{k+\alpha}$; the function $\phi_n$ survives only for integer valued $\theta = n$. In this way we arrive at a finite set of *isolated* eigenvalues $1, 2^{-1}, \cdots, 2^{-k}$, 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 $\phi$ 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 (21.6), 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 $\partial D$ boundary contour:

$$\mathcal{L}_s h(y) = \sigma \oint_{\partial D} \frac{dw}{2\pi i} \frac{h(w) F_s'(y)}{w - F_s(y)} \,. \tag{21.22}$$

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 $s = 1$, 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 $y \in \mathrm{Cl}\, D$. The problem is now reduced to the standard theory for Fredholm determinants, sect. 21.3. 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:

$$\mathrm{tr}\, \mathcal{L}_F = \oint \frac{dw}{2\pi i} \frac{\sigma F'(w)}{w - F(w)} \,.$$

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

[exercise 21.6]

$$\mathrm{tr}\, \mathcal{L}_F = \frac{\sigma F'(x^*)}{1 - F'(x^*)} = \frac{1}{|f'(x^*) - 1|} \,,$$

justifying our previous *ad hoc* calculations of traces using Dirac delta functions.

**Example 21.8** ***Perron-Frobenius operator in a matrix representation:*** *As in example 21.1, 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 $\sigma = \sigma(F') = F'/|F'|$*

$$\mathcal{L}\phi(z) = \int dx\, \delta(z - f(x))\, \phi(x) = \sigma\, F'(z)\, \phi(F(z)) \,.$$

*Assume that $F$ is a contraction of the unit disk, i.e.,*

$$|F(z)| < \theta < 1 \quad \textit{and} \quad |F'(z)| < C < \infty \quad \textit{for} \quad |z| < 1\,, \tag{21.23}$$

*and expand $\phi$ in a polynomial basis by means of the Cauchy formula*

$$\phi(z) = \sum_{n\geq 0} z^n \phi_n = \oint \frac{dw}{2\pi i}\,\frac{\phi(w)}{w-z}\,, \quad \phi_n = \oint \frac{dw}{2\pi i}\,\frac{\phi(w)}{w^{n+1}}$$

*Combining this with (21.22), we see that in this basis $\mathcal{L}$ is represented by the matrix*

$$\mathcal{L}\phi(w) = \sum_{m,n} w^m L_{mn}\phi_n\,, \quad L_{mn} = \oint \frac{dw}{2\pi i}\,\frac{\sigma\, F'(w)(F(w))^n}{w^{m+1}}\,. \tag{21.24}$$

*Taking the trace and summing we get:*

$$\mathrm{tr}\ \mathcal{L} = \sum_{n\geq 0} L_{nn} = \oint \frac{dw}{2\pi i}\,\frac{\sigma\, F'(w)}{w - F(w)}\,.$$

*This integral has but one simple pole at the unique fixed point $w^* = F(w^*) = f(w^*)$. Hence*

$$\mathrm{tr}\ \mathcal{L} = \frac{\sigma\, F'(w^*)}{1 - F'(w^*)} = \frac{1}{|f'(w^*) - 1|}\,.$$

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:

[exercise 21.6]

1) the evolution operator is *multiplicative* along the flow,
2) the symbolic dynamics is a *finite subshift*,
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

$$\mathcal{L}^t(y,x) = |\Lambda^t(x)|^\beta \delta\Big(y - f^t(x)\Big)\ ,$$

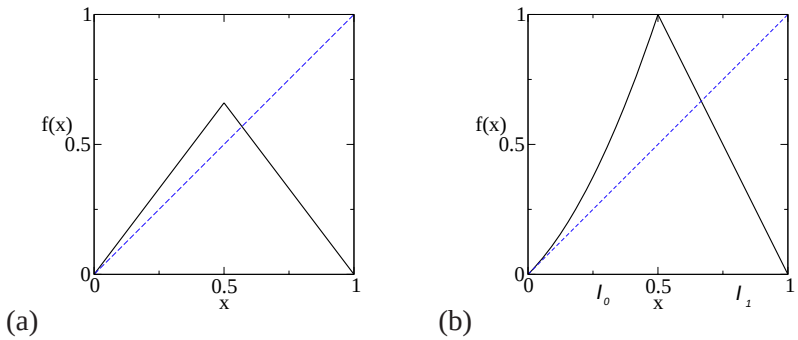



**Figure 21.2:** (a) A (hyperbolic) tent map without a finite Markov partition. (b) A Markov map with a marginal fixed point.

where $\Lambda^t(x)$ is an eigenvalue of the fundamental matrix transverse to the flow. Semiclassical quantum mechanics suggest operators of this form with $\beta = 1/2$.The problem with such operators arises from the fact that when considering the fundamental 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-$d$ tent map (see figure 21.2 (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. 11.5 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 $\alpha$ is varied, the non-leading zeros of the spectral determinant move wildly about.

Property 3 is violated by the map (see figure 21.2 (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 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 23.

Property 4 is required as the heuristic approach of chapter 16 faces two major hurdles:

1. The trace (16.8) 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 of the circle, $x \mapsto 2x \bmod 1$, $x \in R/Z$. 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 length $n$ decreases by 1. The determinant becomes:

$$\det(1 - z\mathcal{L}) = \exp\left(-\sum_{n=1} \frac{z^n}{n} \frac{2^n - 1}{2^n - 1}\right) = 1 - z. \tag{21.25}$$

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.

## 21.5 Hyperbolic maps

> I can give you a definion 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 21.5 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 21.6 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. Her comes an idea; use the analyticity of the map to rewrite the Perron-Frobenius operator acting as follows (where $\sigma$ denotes the sign of the derivative in the unstable direction):

$$\mathcal{L}h(z_1, z_2) = \oint\oint \frac{\sigma\, h(w_1, w_2)}{(z_1 - f_1(w_1, w_2)(f_2(w_1, w_2) - z_2)} \frac{dw_1}{2\pi i} \frac{dw_2}{2\pi i}. \tag{21.26}$$

Here the function $\phi$ 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 (21.9). Such operators form the building blocks in the calculation of traces and determinants. One can prove the following:

**Theorem:** *The spectral determinant for 2-d hyperbolic analytic maps is entire.* [remark 21.8]

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 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 21.8) one may choose coordinate axes close to the stable/unstable manifolds of the map. With the state space divided into $N$ rectangles $\{\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_N\}$, $\mathcal{M}_i = I_i^h \times I_i^v$ 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:

<u>Analytic hyperbolic property:</u> Either $f(\mathcal{M}_i) \cap \mathrm{Int}(\mathcal{M}_j) = \emptyset$, or for each pair $w_h \in \mathrm{Cl}(D_i^h)$, $z_v \in \mathrm{Cl}(D_j^v)$ there exist unique analytic functions of $w_h, z_v$: $w_v = w_v(w_h, z_v) \in \mathrm{Int}(D_i^v)$, $z_h = z_h(w_h, z_v) \in \mathrm{Int}(D_j^h)$, 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 21.3).

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.

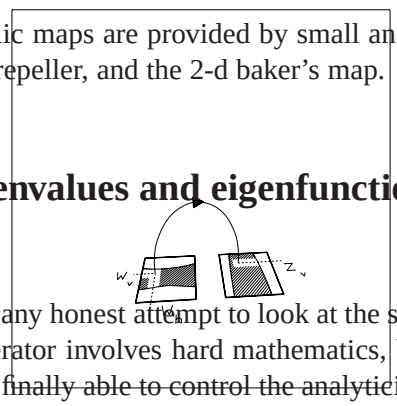



**Figure 21.3:** For an analytic hyperbolic map, specifying the contracting coordinate $w_h$ at the initial rectangle and the expanding coordinate $z_v$ at the image rectangle defines a unique trajectory between the two rectangles. In particular, $w_v$ and $z_h$ (not shown) are uniquely specified.

In two dimensions the operator in (21.26) 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 (16.8) is correct.

Examples of analytic hyperbolic maps are provided by small analytic perturbations of the cat map, the 3-disk repeller, and the 2-d baker's map.

## 21.6 The physics of eigenvalues and eigenfunctions

We appreciate by now that any honest attempt to look at the spectral properties of the Perron-Frobenius operator involves hard mathematics, but the effort is rewarded by the fact that we are ~~finally able to control the analytic~~ity properties of dynamical zeta functions and spectral determinants, and thus substantiate the claim that these objects provide a powerful and well-founded perturbation theory.

Often (see chapter 15) 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 14): 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 the finite dimensional setting, such a theorem is formulated as follows:

[remark 21.7]

- **Perron-Frobenius theorem:** Let $L_{ij}$ be a nonnegative matrix, such that some $n$ exists for which $(L^n)_{ij} > 0$ $\forall i, j$: then
  1. The maximal modulus eigenvalue is non-degenerate real, and positive
  2. The corresponding eigenvector (defined up to a constant) has nonnegative 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 $\theta$. We denote by $\lambda_1, \lambda_2 \ldots, \lambda_M$ the eigenvalues outside of this disk, ordered by the size of their absolute value, with $\lambda_1 = 1$. Then we have the following decomposition

$$\mathcal{L}\varphi = \sum_{i=1}^{M} \lambda_i \psi_i L_i \psi_i^* \varphi + \mathbf{P}\mathcal{L}\varphi \tag{21.27}$$

when $L_i$ are (finite) matrices in Jordan canomical form ($L_0 = 0$ is a $[1\times1]$ matrix, as $\lambda_0$ is simple, due to the Perron-Frobenius theorem), whereas $\psi_i$ is a row vector whose elements form a basis on the eigenspace corresponding to $\lambda_i$, and $\psi_i^*$ 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 $\lambda_i$. For iterates of the Perron-Frobenius operator, (21.27) becomes

$$\mathcal{L}^n\varphi = \sum_{i=1}^{M} \lambda_i^n \psi_i L_i^n \psi_i^* \varphi + \mathbf{P}\mathcal{L}^n\varphi\,. \tag{21.28}$$

If we now consider, for example, correlation between initial $\varphi$ evolved $n$ steps and final $\xi$,

$$\langle\xi|\mathcal{L}^n|\varphi\rangle = \int_{\mathcal{M}} dy\, \xi(y)\,(\mathcal{L}^n\varphi)\,(y) = \int_{\mathcal{M}} dw\,(\xi\circ f^n)(w)\varphi(w)\,, \tag{21.29}$$

it follows that

$$\langle\xi|\mathcal{L}^n|\varphi\rangle = \lambda_1^n\omega_1(\xi,\varphi) + \sum_{i=2}^{L} \lambda_i^n \omega_i^{(n)}(\xi,\varphi) + O(\theta^n)\,, \tag{21.30}$$

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 $\lambda_1$ contribution). Moreover if $\omega_1(\xi,\varphi) = 0$ then (21.29) defines a correlation function: as each term in (21.30) vanishes exponentially in the $n \to \infty$ limit, the eigenvalues $\lambda_2, \ldots, \lambda_M$ determine the exponential decay of correlations for our dynamical system. The prefactors $\omega$ depend on the choice of functions, whereas the exponential decay rates (given by logarithms of $\lambda_i$) 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 21.7]

**Example 21.9** ***Bernoulli shift eigenfunctions:*** *Let us revisit the Bernoulli shift example (21.6) on the space of analytic functions on a disk: apart from the origin we have only simple eigenvalues $\lambda_k = 2^{-k}$, $k = 0, 1, \ldots$. 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 (21.30) the asymptotic decay of the correlation function is (unless also $\omega_1(\eta, \eta) = 0$)*

$$C_{\eta,\eta}(n) \sim \exp(-n \log 2) . \tag{21.31}$$

*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*

$$\eta(z) = \int_0^1 dw\, \eta(w) + \sum_{m=1}^{\infty} \frac{\eta^{(m-1)}(1) - \eta^{(m-1)}(0)}{m!} B_m(z) . \tag{21.32}$$

*As we are considering functions with zero average, we have from (21.29) 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 (21.32) is also useful in realizing that the linear functionals $\psi_i^*$ 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*

$$\Psi_i(w) = \frac{(-1)^{i-1}}{i!} \left( \delta^{(i-1)}(w-1) - \delta^{(i-1)}(w) \right) , \tag{21.33}$$

*when $i \geq 1$ and $\Psi_0(w) = 1$. This representation is only meaningful when the function $\varepsilon$ is analytic in neighborhoods of $w, w-1$.*

## 21.7 Troubles ahead

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:

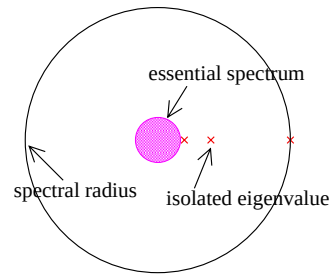



**Figure 21.4:** Spectrum of the Perron-Frobenius operator acting on the space of $\mathbb{C}^{k+\alpha}$ Hölder-continuous functions: only $k$ isolated eigenvalues remain between the spectral radius, and the essential spectral radius which bounds the "essential," continuous spectrum.

- 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 $y^\eta$ with $\mathrm{Re}\,\eta > \mathrm{k}$ is an eigenfunction of the Perron-Frobenius operator and we have [exercise 21.1]

$$\mathcal{L}y^\eta = \frac{1}{|\Lambda|\Lambda^\eta} y^\eta \,, \qquad \eta \in \mathbb{C} \,.$$

  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 21.4. In literature the radius of this disk is called the *essential spectral radius*.

  In sect. 21.4 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 21.2]

- A deceptively innocent assumption is hidden beneath much that was discussed so far: that (21.1) 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. 21.5, extensions of the results obtained for expanding 1-$d$ 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. 21.4.

## Résumé

Examples of analytic eigenfunctions for 1-*d* 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 this book 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 work 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-*d* 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 10 and 11 ) do they have finite grammar. The importance of symbolic dynamics is generally grossly unappreciated; 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 21.1** 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 [34] 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. [1, 2]. An alternative approach to spectral properties of the Perron-Frobenius operator is given in ref. [3].

Ergodic theory, as presented by Sinai [14] 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 [12] and Denker, Grillenberger and Sigmund [16]; and a more formal one is given by Peterson [17].

W. Tucker [28, 29, 30] has proven rigorously via interval arithmetic that the Lorentz attractor is strange for the original parameters, and has a long stable periodic orbit for the slightly different parameters.

**Remark 21.2** Fredholm theory. Our brief summary of Fredholm theory is based on the exposition of ref. [4]. A technical introduction of the theory from an operator point of view is given in ref. [5]. The theory is presented in a more general form in ref. [6].

**Remark 21.3** Bernoulli shift. For a more detailed discussion, consult chapter 3 of ref. [34]. The extension of Fredholm theory to the case or 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 [7]: a concise and readable statement of the results is contained in ref. [8].

**Remark 21.4** 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. [3].

**Remark 21.5** Spectral determinants for smooth flows. The theorem on page 362 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 (3.5) 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 staring at $x$, we reproduce the flow spectral determinant (17.13). Proofs are beyond the scope of this chapter.

**Remark 21.6** Explicit diagonalization. For 1-$d$ repellers a diagonalization of an explicit truncated $L_{mn}$ matrix evaluated in a judiciously chosen basis may yield many more eigenvalues than a cycle expansion (see refs. [10, 11]). The reasons why one persists in using periodic orbit theory are partially aesthetic and partially pragmatic. The explicit calculation of $L_{mn}$ 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 $L_{mn}$ 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 21.7** Perron-Frobenius theorem. A proof of the Perron-Frobenius theorem may be found in ref. [12]. For positive transfer operators, this theorem has been generalized by Ruelle [13].

**Remark 21.8** Axiom A systems. The proofs in sect. 21.5 follow the thesis work of H.H. Rugh [9, 18, 19]. For a mathematical introduction to the subject, consult the excellent review by V. Baladi [1]. It would take us too far afield to give and explain the definition of Axiom A systems (see refs. [23, 24]). Axiom A implies, however, the existence of a Markov partition of the state space from which the properties 2 and 3 assumed on page 350 follow.

**Remark 21.9** Exponential mixing speed of the Bernoulli shift. We see from (21.31) 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.

**Remark 21.10** Left eigenfunctions. We shall never use an explicit form of left eigenfunctions, corresponding to highly singular kernels like (21.33). Many details have been elaborated in a number of papers, such as ref. [20], with a daring physical interpretation.

**Remark 21.11** Ulam's idea. The approximation of Perron-Frobenius operator defined by (14.14) has been shown to reproduce the spectrum for expanding maps, once finer and finer Markov partitions are used [21]. The subtle point of choosing a state space partitioning for a "generic case" is discussed in ref. [22].

## Exercises

21.1. **What space does $\mathcal{L}$ act on?** Show that (21.2) is a complete basis on the space of analytic functions on a disk (and thus that we found the *complete* set of eigenvalues).

21.2. **What space does $\mathcal{L}$ act on?** What can be said about the spectrum of (21.1) on $L^1[0,1]$? Compare the result with figure 21.4.

21.3. **Euler formula.** Derive the Euler formula (21.5)

$$\begin{aligned}\prod_{k=0}^{\infty}(1+tu^k) &= 1+\frac{t}{1-u}+\frac{t^2u}{(1-u)(1-u^2)}+\frac{t^3u^3}{(1-u)(1-u^2)(1-u^3)}\\ &= \sum_{k=0}^{\infty} t^k \frac{u^{\frac{k(k-1)}{2}}}{(1-u)\cdots(1-u^k)}, \qquad |u|<1 \end{aligned} \tag{21.34}$$

21.4. **2-$d$ product expansion****. We conjecture that the expansion corresponding to (21.34) is in this case

$$\begin{aligned}\prod_{k=0}^{\infty}(1+tu^k)^{k+1} &= \sum_{k=0}^{\infty}\frac{F_k(u)}{(1-u)^2(1-u^2)^2\cdots(1-u^k)^2}t^k\\ &= 1+\frac{1}{(1-u)^2}t+\frac{2u}{(1-u)^2(1-u^2)^2}t^2\\ &\quad+\frac{u^2(1+4u+u^2)}{(1-u)^2(1-u^2)^2(1-u^3)^2}t^3+\cdots \end{aligned} \tag{21.35}$$

$F_k(u)$ 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 21.3).

21.5. **Bernoulli shift on $L$ spaces.** Check that the family (21.21) belongs to $L^1([0,1])$. What can be said about the essential spectral radius on $L^2([0,1])$? A useful reference is [24].

21.6. **Cauchy integrals.** Rework all complex analysis steps used in the Bernoulli shift example on analytic functions on a disk.

21.7. **Escape rate.** Consider the escape rate from a strange repeller: find a choice of trial functions $\xi$ and $\varphi$ such that (21.29) 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.

## References


[21.1] V. Baladi, *A brief introduction to dynamical zeta functions*, in: DMV-Seminar 27, *Classical Nonintegrability, Quantum Chaos*, A. Knauf and Ya.G. Sinai (eds), (Birkhäuser, 1997).

[21.2] M. Pollicott, *Periodic orbits and zeta functions*, 1999 AMS Summer Institute on *Smooth ergodic theory and applications*, Seattle (1999), To appear *Proc. Symposia Pure Applied Math., AMS*.

[21.3] M. Viana, *Stochastic dynamics of deterministic systems*, (Col. Bras. de Matemática, Rio de Janeiro,1997)

[21.4] A.N. Kolmogorov and S.V. Fomin, *Elements of the theory of functions and functional analysis* (Dover,1999).

[21.5] R.G. Douglas, *Banach algebra techniques in operator theory* (Springer, New York,1998).

[21.6] A. Grothendieck, *La théorie de Fredholm*, *Bull. Soc. Math. France* **84**, 319 (1956).

[21.7] D. Ruelle, "An extension of the theory of Fredholm determinants," *Inst. Hautes Études Sci. Publ. Math.* **72**, 175-193 (1990).

[21.8] V. Baladi, *Dynamical zeta functions*, in B. Branner and P. Hjorth, eds., Proceedings of the NATO ASI *Real and Complex Dynamical Systems* (1993), (Kluwer Academic Publishers, Dordrecht, 1995)

[21.9] D. Ruelle, "Zeta-Functions for Expanding Maps and Anosov Flows," *Inv. Math.* **34**, 231-242 (1976).

[21.10] F. Christiansen, P. Cvitanović and H.H. Rugh, *J. Phys* **A 23**, L713 (1990).

[21.11] D. Alonso, D. MacKernan, P. Gaspard and G. Nicolis, *Phys. Rev.* **E54**, 2474 (1996).

[21.12] P. Walters, *An introduction to ergodic theory* (Springer, New York 1982).

[21.13] D. Ruelle, *Commun. Math. Phys.* **9**, 267 (1968).

[21.14] Ya.G. Sinai, *Topics in ergodic theory* (Princeton Univ. Press, Princeton 1994).

[21.15] I. Kornfeld, S. Fomin and Ya. Sinai, *Ergodic Theory* (Springer, New York 1982).

[21.16] M. Denker, C. Grillenberger and K. Sigmund, *Ergodic theory on compact spaces* (Springer Lecture Notes in Math. **470**, 1975).

[21.17] K. Peterson, *Ergodic theory* (Cambridge Univ. Press, Cambridge 1983).

[21.18] D. Fried, "The Zeta functions of Ruelle and Selberg I," *Ann. Scient. Éc. Norm. Sup.* **19**, 491 (1986).

[21.19] H.H. Rugh, "The Correlation Spectrum for Hyperbolic Analytic Maps," *Nonlinearity* **5**, 1237 (1992).

[21.20] H.H. Hasegawa and W.C. Saphir, *Phys. Rev.* **A46**, 7401 (1992).

[21.21] G. Froyland, *Commun. Math. Phys.* **189**, 237 (1997).

[21.22] G. Froyland, "Extracting dynamical behaviour via Markov models," in A. Mees (ed.) *Nonlinear dynamics and statistics: Proceedings Newton Institute, Cambridge 1998* (Birkhäuser, 2000);
math-www.uni-paderborn.de/ froyland.

[21.23] V. Baladi, A. Kitaev, D. Ruelle and S. Semmes, "Sharp determinants and kneading operators for holomorphic maps," IHES preprint (1995).

[21.24] A. Zygmund, *Trigonometric series* (Cambridge Univ. Press, Cambridge 1959).

[21.25] J.D. Crawford and J.R. Cary, *Physica* **D6**, 223 (1983)

[21.26] P. Collet and S. Isola, *Commun.Math.Phys.* **139**, 551 (1991)

[21.27] F. Christiansen, S. Isola, G. Paladin and H.H. Rugh, *J. Phys.* **A 23**, L1301 (1990).

[21.28] W. Tucker, "The Lorentz attractor exists," *C. R. Acad. Sci. Paris Sér. I Math* **328**, 1197 (1999).

[21.29] W. Tucker, "A rigorous ODE solver and Smale's 14th problem," *Found. Comput. Math.* **2**, 53 (2002).

[21.30] M. Viana, "What's new on Lorenz strange attractors?" *Math. Intelligencer* **22**, 6 (2000).

# Chapter 22

# 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.

## 22.1 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 $\mathcal{M}_{s_1 s_2 \ldots s_n}$ of trajectories whose symbol sequence starts with a given set of $n$ symbols $s_1 s_2 \ldots s_n$. We can associate many different quantities to these sets. There are geometric measures such as the volume $V(s_1 s_2 \ldots s_n)$, the area $A(s_1 s_2 \ldots s_n)$ or the length $l(s_1 s_2 \ldots s_n)$ of this set. Or in general we can have some measure $\mu(\mathcal{M}_{s_1 s_2 \ldots s_n}) = \mu(s_1 s_2 \ldots s_n)$ of this set. As we have seen in (20.10) the most important is the natural measure, which is the probability that an ergodic trajectory visits the set $\mu(s_1 s_2 \ldots s_n) = P(s_1 s_2 \ldots s_n)$. The natural measure is additive.

Summed up for all possible symbol sequences of length $n$ it gives the measure of the whole state space:

$$\sum_{s_1 s_2 \ldots s_n} \mu(s_1 s_2 \ldots s_n) = 1 \tag{22.1}$$

expresses probability conservation. Also, summing up for the last symbol we get the measure of a one step shorter sequence

$$\sum_{s_n} \mu(s_1 s_2 \ldots s_n) = \mu(s_1 s_2 \ldots s_{n-1}).$$

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

$$\lambda(s_1 s_2 \ldots s_n) = -\frac{1}{n} \log \mu(s_1 s_2 \ldots s_n). \tag{22.2}$$

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 $\lambda$ 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

$$Z_n(\beta) = \sum_{s_1 s_2 \ldots s_n} \mu^\beta(s_1 s_2 \ldots s_n), \tag{22.3}$$

as they are obviously motivated by Gibbs type partition sums of statistical mechanics. The parameter $\beta$ plays the role of inverse temperature $1/k_B T$ and $E(s_1 s_2 \ldots s_n) = -\log \mu(s1 s_2 \ldots s_n)$ is the energy associated with the microstate labeled by $s_1 s_2 \ldots s_n$ We are tempted also to introduce something analogous with the Free energy. In dynamical systems this is called the Rényi entropy [4] defined by the growth rate of the partition sum

$$K_\beta = \lim_{n\to\infty} \frac{1}{n} \frac{1}{1-\beta} \log\left( \sum_{s_1 s_2 \ldots s_n} \mu^\beta(s_1 s_2 \ldots s_n) \right). \tag{22.4}$$

In the special case $\beta \to 1$ we get Kolmogorov entropy

$$K_1 = \lim_{n\to\infty} \frac{1}{n} \sum_{s_1 s_2 \ldots s_n} -\mu(s_1 s_2 \ldots s_n) \log \mu(s_1 s_2 \ldots s_n),$$

while for $\beta = 0$ we recover the topological entropy

$$h_{top} = K_0 = \lim_{n\to\infty} \frac{1}{n} \log N(n),$$

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 (22.2). 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 $\lambda^*$ which maximize the exponent

$$g(\lambda) - \lambda\beta.$$

The exponent is maximal when the derivative of the exponent vanishes

$$g'(\lambda^*) = \beta. \tag{22.5}$$

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 (22.4) we can now connect the Rényi entropies and $g(\lambda)$

$$(\beta - 1)K_\beta = \lambda^*(\beta)\beta - g(\lambda^*(\beta)). \tag{22.6}$$

Equations (22.5) and (22.6) 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

$$g(\lambda) = \lambda\beta^*(\lambda) - (\beta^*(\lambda) - 1)K_{\beta^*(\lambda)}, \tag{22.7}$$

where now $\beta^*(\lambda)$ can be determined from

$$\frac{d}{d\beta^*}[(\beta^* - 1)K_{\beta^*}] = \lambda. \tag{22.8}$$

Obviously, if we can determine the Rényi entropies we can recover the distribution of probabilities from (22.7) and (22.8).

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 (20.10)

$$\mu(s_1, ..., s_n) \approx \frac{e^{n\gamma}}{|\Lambda_{s_1 s_2 ... s_n}|}. \tag{22.9}$$

The partition sum (22.3) now reads

$$Z_n(\beta) \approx \sum_i \frac{e^{n\beta\gamma}}{|\Lambda_i|^\beta}, \tag{22.10}$$

where the summation goes for periodic orbits of length $n$. We can define the characteristic function

$$\Omega(z, \beta) = \exp\left(-\sum_n \frac{z^n}{n} Z_n(\beta)\right). \tag{22.11}$$

According to (22.4) for large $n$ the partition sum behaves as

$$Z_n(\beta) \sim e^{-n(\beta-1)K_\beta}. \tag{22.12}$$

Substituting this into (22.11) 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 (22.10) into (22.11) and introducing prime and repeated periodic orbits as usual we get

$$\Omega(z, \beta) = \exp\left(-\sum_{p,r} \frac{z^{n_p r} e^{\beta\gamma n_p r}}{r|\Lambda_p^r|^\beta}\right).$$

We can see that the characteristic function is the same as the zeta function we introduced for Lyapunov exponents (G.14) 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

$$P(\beta) = (\beta - 1)K_\beta + \beta\gamma, \tag{22.13}$$

since the leading zero of the zeta function is the pressure. The Rényi entropies $K_\beta$, hence the distribution of the exponents $g(\lambda)$ as well, can be calculated via finding the leading eigenvalue of the operator (G.4).

From (22.13) we can get all the important quantities of the thermodynamic formalism. For $\beta = 0$ we get the topological entropy

$$P(0) = -K_0 = -h_{top}. \tag{22.14}$$

For $\beta = 1$ we get the escape rate

$$P(1) = \gamma. \tag{22.15}$$

Taking the derivative of (22.13) in $\beta = 1$ we get Pesin's formula [1] connecting Kolmogorov entropy and the Lyapunov exponent

$$P'(1) = \overline{\lambda} = K_1 + \gamma. \tag{22.16}$$

[exercise 22.1]

It is important to note that, as always, these formulas are strictly valid for nice hyperbolic systems only. At the end of this Chapter we discuss the important problems we are facing in non-hyperbolic cases.

On figure 22.2 we show a typical pressure and $g(\lambda)$ curve computed for the two scale tent map of Exercise 22.4. 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 $\lambda$ 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: $\overline{\lambda} = g(\overline{\lambda})$ and $1 = g'(\overline{\lambda})$.

Next, we are looking at the distribution of trajectories in real space.

## 22.2 Fractal dimensions

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

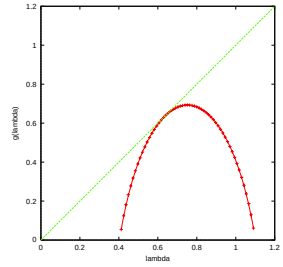



**Figure 22.1**

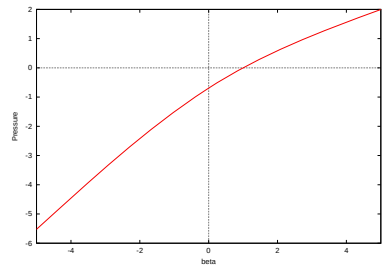



**Figure 22.2:** $g(\lambda)$ and $P(\beta)$ for the map of exercise 22.4 at $a = 3$ and $b = 3/2$. See solution S for details.

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 $\epsilon$ 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 $\alpha_i$ is the local "dimension" or Hölder exponent of the the object. As $\alpha$ 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 $\alpha$ 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 $\alpha$ dependent scaling exponent

$$N_\epsilon(\alpha) \sim \epsilon^{-f(\alpha)}.$$

$f(\alpha)$ can be interpreted [6] as the dimension of the points on the surface of section with scaling exponent $\alpha$. We can calculate $f(\alpha)$ with the help of partition sums as we did for $g(\lambda)$ in the previous section. First, we define

$$Z_\epsilon(q) = \sum_i \mu_i^q. \tag{22.17}$$

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 $\alpha$-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 $\epsilon$ goes to zero the integral is dominated by the term maximizing the exponent. This $\alpha^*$ 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 $\alpha$ 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 [8] $D_q$ via the definition

$$D_q = \tau(q)/(q-1).$$

Some of these dimensions have special names. For $q = 0$ the partition sum (22.17) counts the number of non empty boxes $\bar{N}_\epsilon$. Consequently

$$D_0 = -\lim_{\epsilon\to 0} \frac{\log \bar{N}_\epsilon}{\log \epsilon},$$

is called the box counting dimension. For $q = 1$ the dimension can be determined as the limit of the formulas for $q \to 1$ leading to

$$D_1 = \lim_{\epsilon\to 0} \sum_i \mu_i \log \mu_i / \log \epsilon.$$

This is the scaling exponent of the Shannon information entropy [10] of the distribution, hence its name is *information dimension*.

Using equisize grids is impractical in most of the applications. Instead, we can rewrite (22.17) into the more convenient form

$$\sum_i \frac{\mu_i^q}{\epsilon^\tau(q)} \sim 1. \tag{22.18}$$

If we cover the $i$th branch of the fractal with a grid of size $l_i$ instead of $\epsilon$ we can use the relation [5]

$$\sum_i \frac{\mu_i^q}{l_i{}^\tau(q)} \sim 1, \tag{22.19}$$

the non-uniform grid generalization of 22.18. 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 (22.19) into the form

$$\sum_i \frac{e^{q\gamma n}}{|\Lambda_i|^{q-\tau(q)}} \sim 1, \tag{22.20}$$

where the summation goes for periodic orbits of length $n$. 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 (22.20) 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

$$P(q-(q-1)D_q) = q\gamma. \tag{22.21}$$

Solving this equation directly gives us the partial dimensions of the multifractal repeller along the stable direction. We can see again that the pressure function alone contains all the relevant information. Setting $q = 0$ in (22.21) 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 (22.21) 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

$$D_1 = 1 - \gamma/\overline{\lambda}. \tag{22.22}$$

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 (22.22) proves that the Lyapunov exponent must be larger than the escape rate $\overline{\lambda} > \gamma$ in general.

[exercise 22.4]
[exercise 22.5]
[exercise 22.6]

## 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 $\beta$ 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 22.1** Mild phase transition. In non-hyperbolic systems the formulas derived in this chapter should be modified. As we mentioned in 20.1 in non-hyperbolic systems the periodic orbit expression of the measure can be

$$\mu_0 = e^{\gamma n}/|\Lambda_0|^{\delta} ,$$

where $\delta$ can differ from 1. Usually it is $1/2$. For sufficiently *negative* $\beta$ the corresponding term $1/|\Lambda_0|^{\beta}$ can dominate (22.10) while in (22.3) $e^{\gamma n}/|\Lambda_0|^{\delta\beta}$ plays no dominant role. In this case the pressure as a function of $\beta$ 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 (22.10).

**Remark 22.2** Hard phase transition. The really deep trouble of thermodynamics is caused by intermittency. In that case we have periodic orbits with $|\Lambda_0| \to 1$ as $n \to \infty$. Then for $\beta > 1$ the contribution of these orbits dominate both (22.10) and (22.3). Consequently the partition sum scales as $Z_n(\beta) \to 1$ 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.

## Exercises

22.1. **Thermodynamics in higher dimensions.** Define Lyapunov exponents as the time averages of the eigenexponents of the fundamental matrix $J$

$$\mu^{(k)} = \lim_{t\to\infty} \frac{1}{t} \log|\Lambda_k^t(x_0)|, \tag{22.23}$$

as a generalization of (15.32).

Show that in $d$ dimensions Pesin's formula is

$$K_1 = \sum_{k=1}^{d} \mu^{(k)} - \gamma, \tag{22.24}$$

where the summation goes for the positive $\mu^{(k)}$-s only. Hint: Use the $d$-dimensional generalization of (22.9)

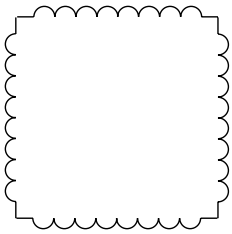
$$\mu_p = e^{n\gamma}/|\prod_k \Lambda_{p,k}|,$$

where the product goes for the expanding eigenvalues of the fundamental matrix of $p$-cycle. (G. Vattay)

22.2. **Stadium billiard Kolmogorov entropy.** (Continuation of exercise 8.4.) Take $a = 1.6$ and $d = 1$ in the stadium billiard figure 8.1, and estimate the Lyapunov exponent by averaging over a very long trajec-

tory. Biham and Kvale [14] 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) $h \approx 1.15 \pm .03$.

22.3. **Entropy of rugged-edge billiards.** Take a semi-circle of diameter $\varepsilon$ and replace the sides of a unit square by $\lfloor 1/\varepsilon \rfloor$ semi-circle arcs.



(a) Is the billiard ergodic as $\varepsilon \to 0$?

(b) (hard) Show that the entropy of the billiard map is

$$K_1 \to -\frac{2}{\pi} \ln \varepsilon + \text{const},$$

as $\varepsilon \to 0$. (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

$$K_1 \to \frac{2 \ln L}{\pi L}.$$

22.4. **Two scale map** Compute all those quantities - dimensions, escape rate, entropies, etc. - for the repeller of the one dimensional map

$$f(x) = \begin{cases} 1 + ax & \text{if} \quad x < 0, \\ 1 - bx & \text{if} \quad x > 0. \end{cases} \tag{22.25}$$

where $a$ and $b$ are larger than 2. Compute the fractal dimension, plot the pressure and compute the $f(\alpha)$ spectrum of singularities.

22.5. **Four scale map** Compute the Rényi entropies and $g(\lambda)$ for the four scale map

$$f(x) = \begin{cases} a_1 x & \text{i} \\ (1-b)((x-b/a_1)/(b-b/a_1)) + b & \text{i} \\ a_2(x-b) & \text{i} \\ (1-b)((x-b-b/a_2)/(1-b-b/a_2)) + b & \text{i} \end{cases}$$

Hint: Calculate the pressure function and use (22.13).

22.6. **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 $I$.

## References


[22.1] J. Balatoni and A. Renyi, *Publi. Math. Inst. Hung. Acad.Sci.* **1**, 9 (1956); (english translation **1**, 588 (Akademia Budapest, 1976)).

[22.2] Ya.B. Pesin, *Uspekhi Mat. Nauk* **32,** 55 (1977), [*Russian Math. Surveys* **32,** 55 (1977)].

[22.3] Even though the thermodynamic formalism is of older vintage (we refer the reader to ref. [4] for a comprehensive overview), we adhere here to the notational conventions of ref. [5] which are more current in the physics literature: we strongly recommend also ref. [6], dealing with period doubling universality.

[22.4] D. Ruelle, *Statistical Mechanics, Thermodynamic Formalism*, (Addison-Wesley, Reading MA, 1978)

[22.5] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia and B.I. Shraiman, *Phys. Rev.* **A107**, 1141 (1986).

[22.6] E. B. Vul, Ya. G. Sinai, and K. M. Khanin, *Uspekhi Mat. Nauk.* **39**, 3 (1984).

[22.7] C. Shannon, *Bell System Technical Journal*, **27**, 379 (1948).

[22.8] V.I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics*, (Addison-Wesley, Redwood City 1989)

[22.9] Ya.G. Sinai, *Topics in Ergodic Theory*, (Princeton University Press, Princeton, New Jersey, 1994)

[22.10] A.N. Kolmogorov, *Dokl.Akad.Nauk.* **124**, 754 (1959)

[22.11] V.I. Arnold, *Mathematical Methods in Classical Mechanics* (Springer-Verlag, Berlin, 1978).

[22.12] C.M. Bender and S.A. Orszag S.A, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw–Hill, Singapore 1978)

[22.13] J.-P. Eckmann and D. Ruelle, *Rev. Mod. Phys.* **57**, 617

[22.14] O. Biham and M. Kvale, *Phys. Rev.* **A 46**, 6334 (1992).

# Chapter 23

# 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 $\{z_0, z_1, z_2, \ldots\}$ 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 phase space, and can be treated by classical methods. In general, one is able to treat the dynamics near stable orbits as well as chaotic components of the phase 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.

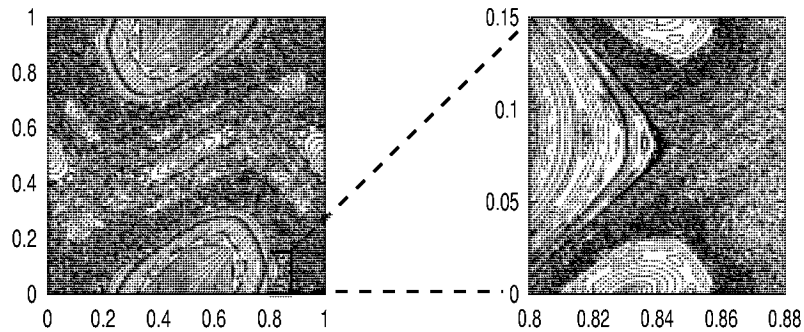



**Figure 23.1:** Typical phase space for an area-preserving map with mixed phase space dynamics; here the standard map for $k = 1.2$ .

## 23.1 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 non-hyperbolic, marginally stable state space regions. Complete hyperbolicity assumed in (16.5) 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 23.1. 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 phase 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 23.1.

Intermittency is due to the existence of fixed points and cycles of marginal stability (5.5), 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 13.2). 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 11, and the intermittency discussed here are two sides of the same coin. [section 11.5]

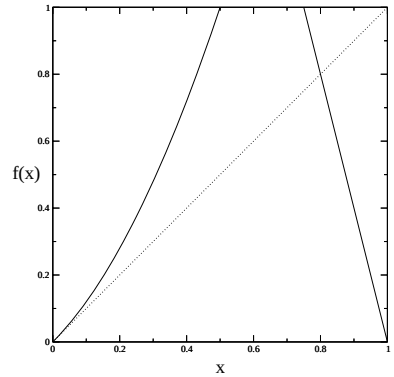



**Figure 23.2:** A complete binary repeller with a marginal fixed point.

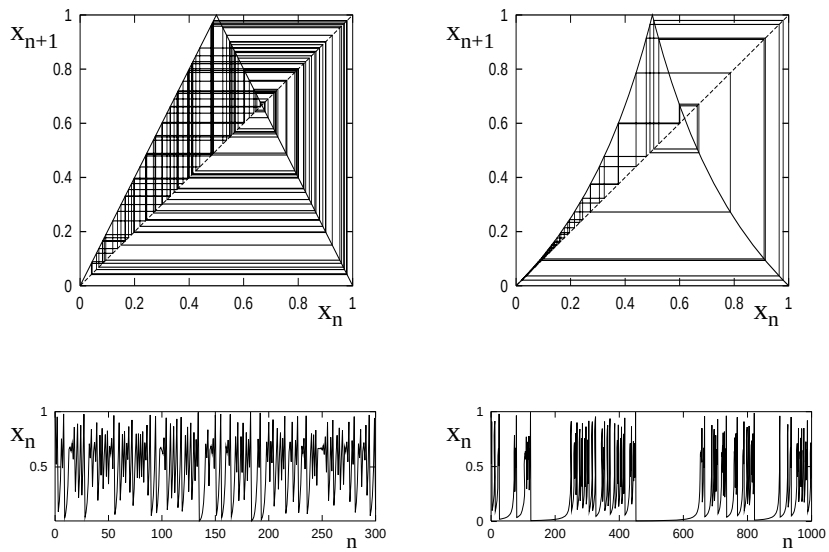



**Figure 23.3:** (a) A tent map trajectory. (b) A Farey map trajectory.

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

$$x \mapsto f(x) = x + O(x^{1+s}) , \tag{23.1}$$

and are expanding everywhere else. Such a map may allow for escape, like the map shown in figure 23.2 or the dynamics may be bounded, like the Farey map (18.31) 163,164c153,154

$$x \mapsto f(x) = \begin{cases} x/(1-x) & x \in [0, 1/2[ \\ (1-x)/x & x \in [1/2, 1] \end{cases}$$

introduced in sect. 18.5.

Figure 23.3 compares a trajectory of the tent map (10.6) 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 18.5.3]

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 23.2 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 21.5 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 (16.3) or the spectral determinant (17.3) 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. 23.2 with a piecewise linear map, with the asymptotics (23.1). 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 15 and 20, 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. 23.2.4.

Branch cuts are typical also for smooth intermittent maps with isolated marginally stable fixed points and cycles. In sect. 23.3, 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. 23.3.1.

Finally, in sect. 23.4, we discuss a probabilistic approach to intermittency that yields approximate dynamical zeta functions and provides valuable information about more complicated systems, such as billiards.

## 23.2 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

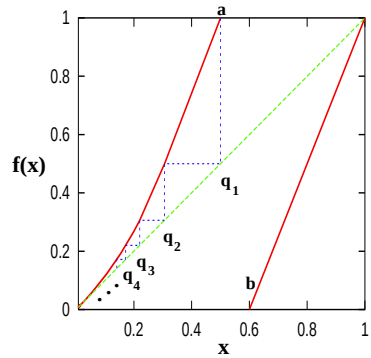



**Figure 23.4:** A piecewise linear intermittent map of (23.2) type: more specifically, the map piecewise linear over intervals (23.8) of the toy example studied below, $a = .5$, $b = .6$, $s = 1.0$.

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. 23.3.

### 23.2.1 A toy map

The Bernoulli shift map (21.6) 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

$$f(x) = \begin{cases} f_0(x) & \text{for } x \in \mathcal{M}_0 = [0, a] \\ f_1(x) & \text{for } x \in \mathcal{M}_1 = [b, 1] \end{cases} . \tag{23.2}$$

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 23.2 and figure 23.4). 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 (23.1) near the origin. We chose a monotonically decreasing sequence of points $q_n$ in $[0, a]$ with $q_1 = a$ and $q_n \to 0$ as $n \to \infty$. 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

$$\mathcal{M}_n = ]q_n, q_{n-1}] \qquad \text{and} \qquad \mathcal{M}_0 = \bigcup_{n=2}^{\infty} \mathcal{M}_n . \tag{23.3}$$

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

$$\begin{array}{rclcll} f_0'(x) & = & \frac{f_0(q_{n-1})-f_0(q_n)}{q_{n-1}-q_n} & = & \frac{|\mathcal{M}_{n-1}|}{|\mathcal{M}_n|} & \text{for } x \in \mathcal{M}_n,\ n \geq 3 \\ f_0'(x) & = & \frac{f_0(q_1)-f_0(q_2)}{q_1-q_2} & = & \frac{1-a}{|\mathcal{M}_2|} & \text{for } x \in \mathcal{M}_2 \\ f_0'(x) & = & \frac{1}{1-b} & = & \frac{|\mathcal{M}|}{|\mathcal{M}_1|} & \text{for } x \in \mathcal{M}_1 \end{array} \tag{23.4}$$

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 (23.1). An orbit $10^n$ enters the intervals $\mathcal{M}_1 \to \mathcal{M}_{n+1} \to \mathcal{M}_n \to \ldots \to \mathcal{M}_2$ successively and the family approaches the marginal stable fixed point at $x = 0$ for $n \to \infty$. The stability of a cycle $10^n$ for $n \geq 1$ is given by the chain rule (4.50),

$$\Lambda_{10^n} = f_0'(x_{n+1}) f_0'(x_n) \ldots f_0'(x_2) f_1'(x_1) = \frac{1}{|\mathcal{M}_{n+1}|} \frac{1-a}{1-b}, \tag{23.5}$$

with $x_i \in \mathcal{M}_i$.

The properties of the map (23.2) are completely determined by the sequence $\{q_n\}$. By choosing $q_n = 2^{-n}$, for example, we recover the uniformly hyperbolic Bernoulli shift map (21.6). An intermittent map of the form (23.3) having the asymptotic behavior (23.1) can be constructed by choosing an algebraically decaying sequence $\{q_n\}$ behaving asymptotically like

$$q_n \sim \frac{1}{n^{1/s}}, \tag{23.6}$$

where $s$ is the intermittency exponent in (23.1). Such a partition leads to intervals whose length decreases asymptotically like a power-law, that is,

$$|\mathcal{M}_n| \sim \frac{1}{n^{1+1/s}} . \tag{23.7}$$

As can be seen from (23.5), the stability eigenvalues of periodic orbit families approaching the marginal fixed point, such as the $10^n$ family increase in turn only algebraically with the cycle length.

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 (23.7) 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 (18.31), and the anomalous diffusion of sect. 24.3. We will, however, not go along 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 [remark 24.8]

$$|\mathcal{M}_n| = C\,\frac{\Gamma(n+m-1/s-1)}{\Gamma(n+m)} \qquad \text{for} \quad n \geq 2, \tag{23.8}$$

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,

$$C = a\left[\sum_{n=m+1}^{\infty} \frac{\Gamma(n-1/s)}{\Gamma(n+1)}\right]^{-1}. \tag{23.9}$$

Using Stirling's formula for the Gamma function

$$\Gamma(z) \sim e^{-z} z^{z-1/2}\,\sqrt{2\pi}\,(1 + 1/12z + \ldots)\,,$$

we verify that the intervals decay asymptotically like $n^{-(1+1/s)}$, as required by the condition (23.7).

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. 18.5 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 (23.5), due to the marginal stability of the fixed point $\overline{0}$. 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

$$10^{n-1} \to n.$$

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 $1, 10, 100, \ldots$:

$$
\begin{aligned}
1/\zeta(z) &= \prod_{p\neq 0}\left(1-\frac{z^{n_p}}{|\Lambda_p|}\right) = 1-\sum_{n=1}^{\infty}\frac{z^n}{|\Lambda_{10^{n-1}}|} \\
&= 1-(1-b)z - C\frac{1-b}{1-a}\sum_{n=2}^{\infty}\frac{\Gamma(n+m-1/s-1)}{\Gamma(n+m)}z^n\,. \qquad (23.10)
\end{aligned}
$$

The fundamental term (18.7) 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 (23.10). Note also that we omitted the fixed point $\overline{0}$ in the above Euler product; we will discussed this point as well as a proper derivation of the zeta function in more detail in sect. 23.2.4.

### 23.2.2 Branch cuts

Starting from the dynamical zeta function (23.10), 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 (23.8). The sum over ratios of Gamma functions in (23.10) can be evaluated analytically by using the following identities valid for $1/s = \alpha > 0$ (the famed binomial theorem in disguise),

- $\alpha$ non-integer

$$
(1-z)^{\alpha} = \sum_{n=0}^{\infty}\frac{\Gamma(n-\alpha)}{\Gamma(-\alpha)\Gamma(n+1)}z^n \qquad (23.11)
$$

- $\alpha$ integer

$$
\begin{aligned}
(1-z)^{\alpha}\log(1-z) &= \sum_{n=1}^{\alpha}(-1)^n c_n z^n \qquad (23.12) \\
&+ (-1)^{\alpha+1}\alpha!\sum_{n=\alpha+1}^{\infty}\frac{(n-\alpha-1)!}{n!}z^n
\end{aligned}
$$

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 (23.11) and (23.12). The infinite sum in (23.10) can now be evaluated with the help of (23.11) or (23.12), that is,

$$\sum_{n=2}^{\infty} \frac{\Gamma(n-1/s)}{\Gamma(n+1)} z^n = \begin{cases} \Gamma(-\frac{1}{s})\left[(1-z)^{1/s} - 1 + \frac{1}{s}z\right] & \text{for} \quad 1 < 1/s < 2; \\ (1-z)\log(1-z) + z & \text{for} \quad s = 1\,. \end{cases}$$

The normalization constant $C$ in (23.8) can be evaluated explicitly using (23.9) and the dynamical zeta function can be given in closed form. We obtain for $1 < 1/s < 2$

$$1/\zeta(z) = 1 - (1-b)z - \frac{a}{1/s - 1}\frac{1-b}{1-a}\left((1-z)^{1/s} - 1 + \frac{1}{s}z\right). \tag{23.13}$$

and for $s = 1$,

$$1/\zeta(z) = 1 - (1-b)z - a\frac{1-b}{1-a}\left((1-z)\log(1-z) + z\right). \tag{23.14}$$

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 $0^n1$ 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 $\mathrm{Re}\, z > 1$ from above and below. The dynamical zeta function for general $s > 0$ takes on the form

$$1/\zeta(z) = 1 - (1-b)z - \frac{a}{g_s(1)}\frac{1-b}{1-a}\frac{1}{z^{m-1}}\left((1-z)^{1/s} - g_s(z)\right) \tag{23.15}$$

for non-integer $s$ with $m = [1/s]$ and

$$1/\zeta(z) = 1-(1-b)z-\frac{a}{g_m(1)}\frac{1-b}{1-a}\frac{1}{z^{m-1}}\left((1-z)^m\log(1-z) - g_m(z)\right) \tag{23.16}$$

for $1/s = m$ integer and $g_s(z)$ are polynomials of order $m = [1/s]$ which can be deduced from (23.11) or (23.12). 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. 23.3 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 $a < b$.

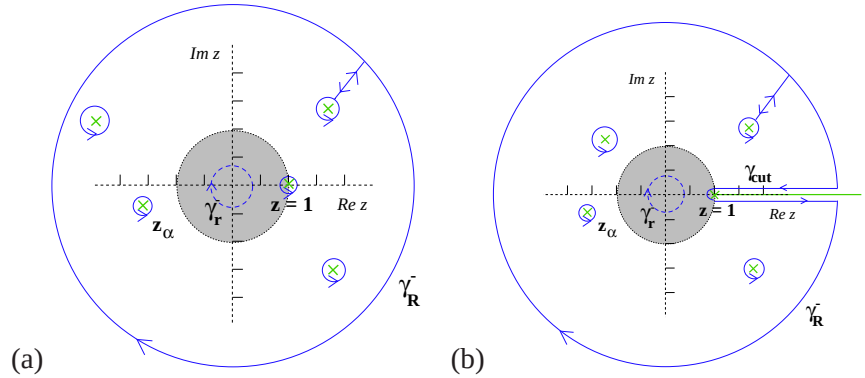



**Figure 23.5:** The survival probability $\Gamma_n$ calculated by contour integration; integrating (23.17) inside the domain of convergence $|z| < 1$ (shaded area) of $1/\zeta(z)$ in periodic orbit representation yields (16.26). A deformation of the contour $\gamma_r^-$ (dashed line) to a larger circle $\gamma_R^-$ gives contributions from the poles and zeros (x) 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 $\gamma_{cut}$ running along the branch cut (b).

### 23.2.3 Escape rate

Our starting point for the calculation of the fraction of survivors after $n$ time steps, is the integral representation (17.19)

$$\Gamma_n = \frac{1}{2\pi i} \oint_{\gamma_r^-} z^{-n} \left( \frac{d}{dz} \log \zeta^{-1}(z) \right) dz \,, \tag{23.17}$$

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 (16.26) 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 23.5 (a). Residue calculus turns this into a sum over the zeros $z_\alpha$ and poles $z_\beta$ of the dynamical zeta function, that is

$$\Gamma_n = \sum_{|z_\alpha|<R}^{\text{zeros}} \frac{1}{z_\alpha^n} - \sum_{|z_\beta|<R}^{\text{poles}} \frac{1}{z_\beta^n} + \frac{1}{2\pi i} \oint_{\gamma_R^-} dz\, z^{-n} \frac{d}{dz} \log \zeta^{-1}, \tag{23.18}$$

where the last term gives a contribution from a large circle $\gamma_R^-$. We thus find exponential decay of $\Gamma_n$ dominated by the leading zero or pole of $\zeta^{-1}(z)$.

Things change considerably in the intermittent case. The point $z = 1$ is a branch cut singularity and there exists no Taylor series expansion of $\zeta^{-1}$ around $z = 1$. Second, the path deformation that led us to (23.18) 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 $\mathrm{Re}\,(z) \geq 1$, $\mathrm{Im}\,(z) = 0$ encircling the branch cut in anti-clockwise direction, see figure 23.5 (b). We will denote the detour around the cut as $\gamma_{cut}$. 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 (23.17) at the branch point $z = 1$. We restrict ourselves again to the case $1 < 1/s < 2$ first and write the dynamical zeta function (23.13) 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

$$\frac{1}{2\pi i}\oint_{\gamma_{cut}} (1-u)^{-n}\frac{d}{du}\log G(u)du \tag{23.19}$$

where $\gamma_{cut}$ 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

$$\frac{d}{du}\log G(u) = \frac{a_1}{a_0} + \frac{1}{s}\frac{b_0}{a_0}u^{1/s-1} + O(u)\,. \tag{23.20}$$

The integrals along the cut may be evaluated using the general formula

$$\frac{1}{2\pi i}\oint_{\gamma_{cut}} u^{\alpha}(1-u)^{-n}du = \frac{\Gamma(n-\alpha-1)}{\Gamma(n)\Gamma(-\alpha)} \sim \frac{1}{n^{\alpha+1}}(1+O(1/n)) \tag{23.21}$$

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 $u^{\alpha}$ at $u = 1$, cf. (23.11). For the continuous time case the corresponding formula is

$$\frac{1}{2\pi i}\oint_{\gamma_{cut}} z^{\alpha}e^{zt}dz = \frac{1}{\Gamma(-\alpha)}\frac{1}{t^{\alpha+1}}. \tag{23.22}$$

Plugging (23.20) into (23.19) and using (23.21) we get the asymptotic result

$$\Gamma_n \sim \frac{b_0}{a_0}\frac{1}{s}\frac{1}{\Gamma(1-1/s)}\frac{1}{n^{1/s}} = \frac{a}{s-1}\frac{1-b}{b-a}\frac{1}{\Gamma(1-1/s)}\frac{1}{n^{1/s}}\,. \tag{23.23}$$

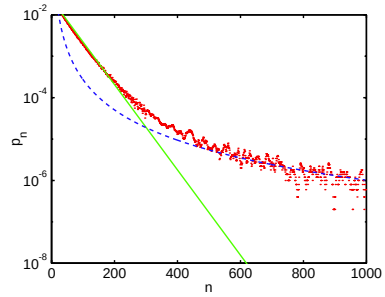



**Figure 23.6:** The asymptotic escape from an intermittent repeller is a power law. Normally it is preceded by an exponential, which can be related to zeros close to the cut but beyond the branch point $z = 1$, as in figure 23.5 (b).

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 23.6.

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 (23.15). Expanding $\frac{d}{du} \log G(u)$ around $u = 0$, one again obtains leading order contributions according to (23.20) and the general result follows immediately using (23.21), that is,

$$\Gamma_n \sim \frac{a}{s g_s(1)} \frac{1-b}{b-a} \frac{1}{\Gamma(1-1/s)} \frac{1}{n^{1/s}} \,. \tag{23.24}$$

Applying the same arguments for integer intermittency exponents $1/s = m$, one obtains

$$\Gamma_n \sim (-1)^{m+1} \frac{a}{s g_m(1)} \frac{1-b}{b-a} \frac{m!}{n^m} \,. \tag{23.25}$$

So far, we have considered the survival probability for a repeller, that is we assumed $a < b$. The formulas (23.24) and (23.25) 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 $G(u)$ (23.20) 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 $\Gamma_n$ defined as sum over the instabilities of the periodic orbits as in (20.12) 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.

### 23.2.4 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 $\Gamma_n$ given by (1.2), (20.2), 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 (16.26). The only orbit to worry about is the marginal fixed point $\overline{0}$ itself which we excluded from the zeta function (23.10).

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 20 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 23.2).

The generating partition is composed of the domains $\mathcal{M}_s$. The $n$th level partition $C^{(n)} = \{\mathcal{M}_i\}$ can be constructed iteratively. Here $i$'s are words $i = s_2 s_2 \ldots s_n$ of length $n$, and the intervals $\mathcal{M}_i$ are constructed recursively

$$\mathcal{M}_{sj} = f_s^{-1}(\mathcal{M}_j)\,, \tag{23.26}$$

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 $\Gamma_n$, 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 (20.6), the stability $\Lambda_i$ of the periodic orbit $i$ within the $\mathcal{M}_i$, and bound the survival probability by the periodic orbit sum (20.7).

The bound (20.6)

$$C_1 \frac{1}{|\Lambda_i|} < \frac{|\mathcal{M}_i|}{|\mathcal{M}|} < C_2 \frac{1}{|\Lambda_i|}$$

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 17.7]

However, bounding survival probability strip by strip is not what is required for establishing the bound (20.7). 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

$$C_1 \frac{1}{|\Lambda_p|} < \frac{1}{n_p} \sum_{i \in p} \frac{|\mathcal{M}_i|}{|\mathcal{M}|} < C_2 \frac{1}{|\Lambda_p|}, \tag{23.27}$$

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 (20.7).

To study averages of multiplicative weights we follow sect. 15.1 and introduce a state space observable $a(x)$ and the integrated quantity

$$A^n(x) = \sum_{k=0}^{n-1} a(f^k(x)).$$

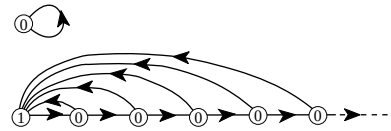



**Figure 23.7:** Markov graph corresponding to the alphabet $\{0^{k-1}1; \overline{0}\,,\ k \geq 1\}$

This leads us to introduce the generating function (15.10)

$$\langle e^{\beta\, A^n(x)} \rangle,$$

where $\langle . \rangle$ 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

$$C_1 \frac{e^{\beta A_p}}{|\Lambda_p|} < \frac{1}{n_p} \sum_{i \in p} \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}_Q} e^{\beta A^n(x)} dx < C_2 \frac{e^{\beta A_p}}{|\Lambda_p|}, \tag{23.28}$$

is valid for all $p$. After performing the above average one gets

$$C_1 \Gamma_n(\beta) < \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} e^{\beta A(x,n)} dx < C_2 \Gamma_n(\beta), \tag{23.29}$$

with

$$\Gamma_n(\beta) = \sum_p^n \frac{e^{\beta A_p}}{|\Lambda_p|}. \tag{23.30}$$

and a dynamical zeta function can be derived. In the intermittent case one can expect that the bound (23.28) holds using an averaging argument similar to the one discussed in (23.27). 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 $\{0^{k-1}1, \overline{0}\,;\ k \geq 1\}$. The symbol 0 occurs unaccompanied by any 1's only in the $\overline{0}$ marginal fixed point which is disconnected from the rest of the Markov graph see figure 23.7.

[chapter 11]

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 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 21]

In the intermittent case, the fixed point $\overline{0}$ does not provide any shadowing , and a statement such as

$$\Lambda_{1 \cdot 0^{k+1}} \approx \Lambda_{1 \cdot 0^k} \Lambda_0,$$

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 (17.15), 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 $\overline{0}$ cycle is a dire necessity. The cycle weight $\Lambda_{\overline{0}}^n = 1$ overestimates the corresponding interval length of $\mathcal{M}_{\overline{0}^n}$ in the partition of the phase space $\mathcal{M}$ by an increasing amount thus leading to wrong results when calculating escape. By leaving out the $\overline{0}$ cycle (and thus also the $\mathcal{M}_{\overline{0}^n}$ contribution), we are guaranteed to get at least the right asymptotical behavior.

Note also, that if we are working with the spectral determinant (17.3), 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 $\overline{0}$ cycle.

## 23.3 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 23.3, or the repeller depicted in figure 23.2. We again assume that the behavior near the fixed point

**Table 23.1:** Infinite alphabet versus the original binary alphabet for the shortest periodic orbit families. Repetitions of prime cycles ($11 = 1^2, 0101 = 01^2, \ldots$) and their cyclic repeats ($110 = 101, 1110 = 1101, \ldots$) are accounted for by cancelations and combination factors in the cycle expansion (23.31).

| $\infty$ – 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 |

is given by (23.1). 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 $\overline{0}$; periodic orbit contributions of the form $t_{0^n1}$ are now unbalanced and we arrive at a cycle expansion in terms of infinitely many fundamental terms as for our toy map. This corresponds to moving from our binary symbolic dynamics to an infinite symbolic dynamics by making the identification

$$10^{n-1} \to n; \quad 10^{n-1}10^{m-1} \to nm; \quad 10^{n-1}10^{m-1}10^{k-1} \to nmk; \ldots$$

see also table **??**. 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 11.1]

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 $t_{nm\ldots} = t_{10^{n-1}10^{m-1}\ldots}$, one obtains

$$\begin{aligned}\zeta^{-1} &= \prod_{p\neq 0}\left(1-t_p\right) = 1-\sum_{n=1}^{\infty} ce \\ &= 1-\sum_{n=1}^{\infty} t_n - \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{1}{2}(t_{mn}-t_m t_n) \\ &- \sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}(\frac{1}{3}t_{kmn}-\frac{1}{2}t_{km}t_n+\frac{1}{6}t_k t_m t_n) - \sum_{l=1}^{\infty}\sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\ldots\,.\end{aligned} \tag{23.31}$$

The first sum is the fundamental term, which we have already seen in the toy model, (23.10). 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

$$f(z) = \sum_{n=0}^{\infty} h_n z^n. \tag{23.32}$$

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 (23.11) or (23.12), one can in general construct an analytic continuation of $f(z)$ around $z = 1$ of the form

$$\begin{aligned} f(z) &\sim A(z) + (1-z)^{\alpha-1}B(z) \qquad \alpha \notin \mathbb{N} \\ f(z) &\sim A(z) + (1-z)^{\alpha-1}\ln(1-z) \qquad \alpha \in \mathbb{N}\,, \end{aligned} \tag{23.33}$$

where $A(z)$ and $B(z)$ are functions analytic in a disc around $z = 1$. We thus again find that the zeta function (23.31) has a branch cut along the real axis $\mathrm{Re}\, z \geq 1$. 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 (23.24) and (23.25) is a general result, that is, the survival probability is given asymptotically by

$$\Gamma_n \sim C\frac{1}{n^{1/s}} \tag{23.34}$$

for all 1-dimensional maps of the form (23.1). 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.

### 23.3.1 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 (23.33) at $z = 1$. This can be done in principle, if the coefficients $h_n$ in sums like (23.32) are known (as for our toy model). One then considers a resummation of the form

$$\sum_{j=0}^{\infty} h_j z^j = \sum_{j=0}^{\infty} a_j (1-z)^j + (1-z)^{\alpha-1} \sum_{j=0}^{\infty} b_j (1-z)^j, \tag{23.35}$$

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 (23.11) and equating the coefficients.

In practical calculations one often has only a finite number of coefficients $h_j$, $0 \le j \le N$, 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 (23.35). We replace the infinite sums in (23.35) by finite sums of increasing degrees $n_a$ and $n_b$, and require that

$$\sum_{i=0}^{n_a} a_i (1-z)^i + (1-z)^{\alpha-1} \sum_{i=0}^{n_b} b_i (1-z)^i = \sum_{i=0}^{N} h_i z^i + O(z^{N+1}) \ . \tag{23.36}$$

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 (23.36) 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 (23.34), cf. (23.23). 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 (23.33) holds in a disc of radius 1 around $z = 1$. Convergence is improved further if additional information about the asymptotics of sums like (23.32) is used to improve the ansatz (23.35).

### 23.3.2 Analytical continuation by integral transformations

We will now introduce a method which provides an analytic continuation of sums of the form (23.32) without explicitly relying on an ansatz (23.35). The main idea is to rewrite the sum (23.32) 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

$$h(n) = n^{-\alpha}(C_1 + C_2 n^{-1} + \ldots), \qquad n \neq 0,$$

and determine the constants $C_i$ numerically from periodic orbit data. By using the Poisson resummation identity

$$\sum_{n=-\infty}^{\infty} \delta(x-n) = \sum_{m=-\infty}^{\infty} \exp(2\pi i m x), \tag{23.37}$$

we may write the sum as (23.32)

$$f(z) = \frac{1}{2}h(0) + \sum_{m=-\infty}^{\infty} \int_0^{\infty} dx\, e^{2\pi i m x} h(x) z^x. \tag{23.38}$$

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 (23.38), we transform $x \to ix$ 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 $2\pi$. The last condition reminds us again of the existence of a branch cut at $\mathrm{Re}\, z \geq 1$. By the same technique, we find the analytic continuation for all the other integrals in (23.38). The real axis is then rotated according to $x \to \mathrm{sign}(m) i x$ where $\mathrm{sign}(m)$ refers to the sign of $m$.

$$\int_0^{\infty} dx\, e^{\pm 2\pi i |m| x} h(x)\, r^x e^{ix\sigma} = \pm \mathrm{i} \int_0^{\infty} dx\, h(\pm ix)\, r^{\pm ix} e^{-x(2\pi|m| \pm \sigma)}.$$

Changing summation and integration, we can carry out the sum over $|m|$ explicitly and one finally obtains the compact expression

$$\begin{aligned} f(z) &= \frac{1}{2}h(0) + i\int_0^\infty dx\, h(ix)\, r^{ix} e^{-x\sigma} \\ &+ i\int_0^\infty dx\, \frac{e^{-2\pi x}}{1-e^{-2\pi x}} \left[ h(ix) r^{ix} e^{-x\sigma} - h(-ix) r^{-ix} e^{x\sigma} \right]. \end{aligned} \tag{23.39}$$

The transformation from the original sum to the two integrals in (23.39) is exact for $r \leq 1$, and provides an analytic continuation for $r > 0$. The expression (23.39) 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.

### 23.3.3 Curvature contributions

So far, we have discussed only the fundamental term $\sum_{n=1}^{\infty} t_n$ in (23.31), 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 (23.33) are contained in the curvature terms in (23.31). 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 (23.5) suggests that $t_{mn}$ decrease asymptotically as

$$t_{mn} \sim \frac{1}{(nm)^{1+1/s}} \tag{23.40}$$

for 2-cycles, and in general for $n$-cycles as

$$t_{m_1 m_2 \ldots m_n} \sim \frac{1}{(m_1 m_2 \ldots m_n)^{1+1/s}}\,.$$

If we happen to know the cycle weights $t_{m_1 m_2 \ldots m_n}$ analytically, we may proceed as in sect. 23.3.2, 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.

## 23.4 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 (3.5), 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 cycle point. The dynamical zeta function (17.15)

$$1/\zeta(z,s,\beta) = \prod_p \left(1 - \frac{z^{n_p} e^{\beta A_p - sT_p}}{|\Lambda_p|}\right), \qquad A_p = \sum_{k=0}^{n_p-1} a(f^k(x_p)), \tag{23.41}$$

[chapter 15]

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 $1/\zeta(z,\beta) = 1/\zeta(z,0,\beta)$.

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 $e^{\beta A(x_0,n)-sT(x_0,n)}$ averaged over the surface of section $\mathcal{P}$. With the above probabilistic assumption the large $n$ behavior is

$$\langle e^{\beta A(x_0,n)-sT(x_0,n)}\rangle_{\mathcal{P}} \sim \left(\int_{\mathcal{P}} e^{\beta a(x)-s\tau}\rho(x)dx\right)^n,$$

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 $z$-$\beta$ 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 15.1]

$$1/\zeta(z,s,\beta) = z - \int_{\mathcal{P}} e^{\beta a(x) - s\tau(x)} \rho(x) dx \quad , \tag{23.42}$$

where we have neglected a possible analytic and non-zero prefactor. The dynamical zeta function of the flow is now

$$1/\zeta(s,\beta) = 1/\zeta(1,s,\beta) = 1 - \int_{\mathcal{P}} e^{\beta a(x)} \rho(x) e^{-s\tau(x)} dx \quad . \tag{23.43}$$

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 $1/\zeta_{\rm BER}(s,\beta)$ in what follows.

A central role is played by the probability distribution of return times

$$\psi(\tau) = \int_{\mathcal{P}} \delta(\tau - \tau(x)) \rho(x) dx \tag{23.44}$$

[exercise 24.6]

The BER zeta function at $\beta = 0$ is then given in terms of the Laplace transform of this distribution

$$1/\zeta_{\rm BER}(s) = 1 - \int_0^\infty \psi(\tau) e^{-s\tau} d\tau.$$

[exercise 23.5]

**Example 23.1** ***Return times for the Bernoulli map.*** *For the Bernoulli shift map (21.6)*

$$x \mapsto f(x) = 2x \ mod\ 1,$$

*one easily derives the distribution of return times*

$$\psi_n = \frac{1}{2^n} \qquad n \geq 1.$$

*The BER zeta function becomes (by the discrete Laplace transform (16.9))*

$$\begin{aligned} 1/\zeta_{BER}(z) &= 1 - \sum_{n=1}^{\infty} \psi_n z^n = 1 - \sum_{n=1}^{\infty} \frac{z^n}{2^n} \\ &= \frac{1-z}{1-z/2} = \zeta^{-1}(z)/(1 - z/\Lambda_0) \,. \end{aligned} \tag{23.45}$$

*Thanks to the uniformity of the piecewise linear map measure (15.19) the "approximate" zeta function is in this case the* exact *dynamical zeta function, with the cycle point* $\overline{0}$ *pruned.*

**Example 23.2** ***Return times for the model of sect. 23.2.1.*** *For the toy model of sect. 23.2.1 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*

$$1/\zeta_{BER}(z) = 1 - z|\mathcal{M}_1| - \sum_{n=2}^{\infty} |\mathcal{M}_n| z^n,$$

*which again coincides with the exact result, (23.10).*

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. 18.1.1 are indeed intricately related, but not identical in general. In particular, note that the BER zeta function obeys the flow conservation sum rule (20.11) 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 non-hyperbolic 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 23.1** What about the evolution operator formalism? The main virtue of evolution operators was their semigroup property (15.25). 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 [2]. 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. [3, 20] for detailed discussions of this technique, as well as applications to 1-dimensional maps.

**Remark 23.2** Intermittency. Intermittency was discovered by Manneville and Pomeau [1] in their study of the Lorentz 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. [5] they explained this phenomenon in terms of a 1-dimensional map (such as (23.1)) near tangent bifurcation, and classified possible types of intermittency.

Piecewise linear models like the one considered here have been studied by Gaspard and Wang [6]. The escape problem has here been treated following ref. [7], resummations following ref. [8]. The proof of the bound (23.27) can be found in P. Dahlqvist's notes on `ChaosBook.org/PDahlqvistEscape.ps.gz`.

Farey map (18.31) has been studied widely in the context of intermittent dynamics, for example in refs. [16, 17, 3, 18, 19, 14, 2]. The Fredholm determinant and the dynamical zeta functions for the Farey map (18.31) and the related Gauss shift map (14.46) have been studied by Mayer [16]. 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 [24] for "Axiom A" hyperbolic systems.

**Remark 23.3** Tauberian theorems. In this chapter we used Tauberian theorems for power series and Laplace transforms: Feller's monograph [9] is a highly recommended introduction to these methods.

**Remark 23.4** Probabilistic methods, BER zeta functions. Probabilistic description of intermittent chaos was introduced by Geisal and Thomae [10]. The BER approximation studied here is inspired by Baladi, Eckmann and Ruelle [14], with further developments in refs. [13, 15].

## Exercises

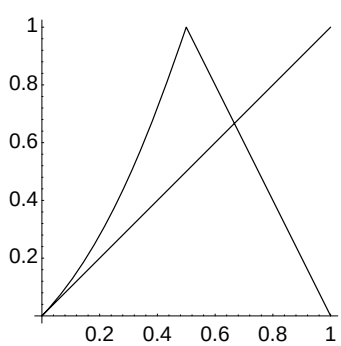
23.1. **Integral representation of Jonquière functions.** Check the integral representation

$$J(z,\alpha) = \frac{z}{\Gamma(\alpha)} \int_0^\infty d\xi \, \frac{\xi^{\alpha-1}}{e^\xi - z} \qquad \text{for} \quad \alpha > 0 \,. \tag{23.46}$$

Note how the denominator is connected to Bose-Einstein distribution. Compute $J(x+i\epsilon) - J(x-i\epsilon)$ for a real $x > 1$.

23.2. **Power law correction to a power law.** Expand (23.20) further and derive the leading power law correction to (23.23).

23.3. **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 $f \to x$ as $x \to 0$. 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

$$\Lambda_{\underbrace{00\cdots0}_{n}1} \sim n^2$$

23.4. **Power law fall-off of stability eigenvalues 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 (8.13) is that the $\Lambda_n$ grows only as a power law in number of bounces:

$$\Lambda_n \propto n_k^2 \,. \tag{23.47}$$

Check.

23.5. **Probabilistic zeta function for maps.** Derive the probabilistic zeta function for a map with recurrence distribution $\psi_n$.

23.6. **Accelerated diffusion.** Consider a map $h$, such that $\hat{h} = \hat{f}$, but now running branches are turner 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

$$\sigma^2(t) \sim \begin{cases} t & \text{for } \alpha > 2 \\ t \ln t & \text{for } \alpha = 2 \\ t^{3-\alpha} & \text{for } \alpha \in (1,2) \\ t^2/\ln t & \text{for } \alpha = 1 \\ t^2 & \text{for } \alpha \in (0,1) \end{cases}$$

23.7. **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 the diffusion constant (15.13) 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

$$f_\epsilon(x) = \begin{cases} \Lambda x & \text{for } x \in \left[0, x_1^+\right] \\ a - \Lambda_{\epsilon,\gamma}|x - x^+| & \text{for } x \in \left[x_1^+, x_2^+\right] \\ 1 - \Lambda'(x - x_2^+) & \text{for } x \in \left[x_2^+, x_1^-\right] \\ 1 - a + \Lambda_{\epsilon,\gamma}|x - x^-| & \text{for } x \in \left[x_1^-, x_2^-\right] \\ 1 + \Lambda(x-1) & \text{for } x \in \left[x_2^-, 1\right] \end{cases} \tag{23.}$$

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 (24.11) 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 $\gamma$ 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

$$\{1, 2, 3, s_\# \cdot 1^i, d_\# \cdot 3^k\} \qquad \# = a, b \quad i, k = m, m+1, m$$

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 (24.8) we get

$$D = \frac{2\epsilon_m^{1/\gamma-1}\Lambda^{-m}(1-1/\Lambda)^{-1}}{1-\frac{2}{\Lambda}-\frac{1}{\Lambda'}-4\epsilon_m^{1/\gamma-1}\left(\frac{m+1}{\Lambda^m(1-1/\Lambda)}+\frac{1}{\Lambda^{m+1}(1-1/\Lambda)^2}\right)} \tag{23.49}$$

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 (23.49) 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 23.5** Onset of diffusion for continuous maps. The zoology of behavior for continuous maps at the onset of diffusion is described in refs. [12, 13, 25]: our treatment for piecewise linear maps was introduced in ref. [26].

## References


[23.1] P. Manneville and Y. Pomeau, *Phys. Lett.* **75A**, 1 (1979).

[23.2] H.H. Rugh, *Inv. Math.* **135**, 1 (1999).

[23.3] T. Prellberg, *Maps of the interval with indifferent fixed points: thermodynamic formalism and phase transitions*, Ph.D. Thesis, Virginia Polytechnic Institute (1991); T. Prellberg and J. Slawny, "Maps of intervals with indifferent fixed points - thermodynamic formalism and phase transitions," *J. Stat. Phys.* **66**, 503 (1992).

[23.4] T. Prellberg, *Towards a complete determination of the spectrum of a transfer operator associated with intermittency*, *J. Phys. A* **36**, 2455 (2003).

[23.5] Y. Pomeau and P. Manneville, *Commun. Math. Phys.* **74**, 189 (1980).

[23.6] P. Gaspard and X.-J. Wang, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4591 (1988); X.-J. Wang, *Phys. Rev.* **A40**, 6647 (1989); X.-J. Wang, *Phys. Rev.* **A39**, 3214 (1989).

[23.7] P. Dahlqvist, *Phys. Rev. E* **60**, 6639 (1999).

[23.8] P. Dahlqvist, *J. Phys. A* **30**, L351 (1997).

[23.9] W. Feller, *An introduction to probability theory and applications, Vol. II* (Wiley, New York 1966).

[23.10] T. Geisel and S. Thomae, *Phys. Rev. Lett.* **52**, 1936 (1984).

[23.11] T. Geisel, J. Nierwetberg and A. Zacherl, *Phys. Rev. Lett.* **54**, 616 (1985).

[23.12] R. Artuso, G. Casati and R. Lombardi, *Phys. Rev. Lett.* **71**, 62 (1993).

[23.13] P. Dahlqvist, *Nonlinearity* **8**, 11 (1995).

[23.14] V. Baladi, J.-P. Eckmann and D. Ruelle, *Nonlinearity* **2**, 119 (1989).

[23.15] P. Dahlqvist, *J. Phys. A* **27**, 763 (1994).

[23.16] D.H. Mayer, *Bull. Soc. Math. France* **104**, 195 (1976).

[23.17] D. Mayer and G. Roepstorff, *J. Stat. Phys.* **47**, 149 (1987).

[23.18] D. H. Mayer, *Continued fractions and related transformations*, in ref. [2].

[23.19] D. H. Mayer, *The Ruelle-Araki transfer operator in classical statistical mechanics* (Springer-Verlag, Berlin, 1980).

[23.20] S. Isola, *J. Stat. Phys.* **97**, 263 (1999).

[23.21] S. Isola, "On the spectrum of Farey and Gauss maps," `mp-arc 01-280`.

[23.22] B. Fornberg and K.S. Kölbig, *Math. of Computation* **29**, 582 (1975).

[23.23] A. Erdélyi, W. Magnus, F. Oberhettinger and F. G. Tricomi, *Higher transcendental functions, Vol. I* (McGraw-Hill, New York, 1953).

[23.24] D. Ruelle, *Inventiones math.* **34**, 231 (1976)

[23.25] S. Grossmann and H. Fujisaka, *Phys. Rev. A* **26**, 1779 (1982).

[23.26] R. Lombardi, Laurea thesis, Universitá degli studi di Milano (1993).

# 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. 14.1 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 for 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-$d$ 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-$d$ periodic Lorentz gas. This system, however, is not the best

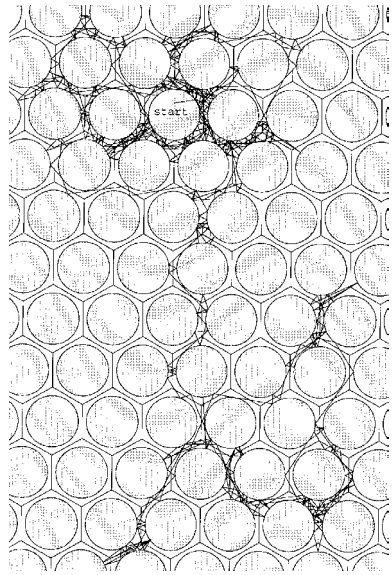



**Figure 24.1:** Deterministic diffusion in a finite horizon periodic Lorentz gas. (T. Schreiber)

one to exemplify the theory, due to its complicated symbolic dynamics. Therefore we apply the theory first to diffusion induced by a 1-*d* maps in sect. 24.2.

## 24.1 Diffusion in periodic arrays

The 2-*d* *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 nontrivial 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. If the scattering array has further discrete symmetries, such as reflection symmetry, each elementary cell may be built from a *fundamental domain* $\widetilde{\mathcal{M}}$ 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.

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 24.2. We distinguish two types

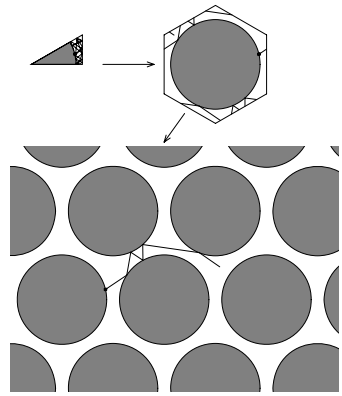


**Figure 24.2:** Tiling of $\hat{\mathcal{M}}$, a periodic lattice of reflecting disks, by the fundamental domain $\widetilde{\mathcal{M}}$. Indicated is an example of a global trajectory $\hat{x}(t)$ together with the corresponding elementary cell trajectory $x(t)$ and the fundamental domain trajectory $\tilde{x}(t)$. (Courtesy of J.-P. Eckmann)

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. 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 return to 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 hats, tildes and nothings atop symbols signify:

| | | |
|---|---|---|
| $\tilde{}$ | fundamental domain, triangle in figure 24.2 | |
| | elementary cell, hexagon in figure 24.2 | |
| $\hat{}$ | full state space, lattice in figure 24.2 | (24.1) |

It is convenient to define an evolution operator for each of the 3 cases of figure 24.2. $\hat{x}(t) = \hat{f}^t(\hat{x})$ 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

$$\hat{n}_t(x_0) = \hat{f}^t(x_0) - f^t(x_0) \in T \,, \tag{24.2}$$

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 $\widetilde{\mathcal{M}}$; $\tilde{f}^t(\tilde{x})$ is related to $f^t(\tilde{x})$ by a discrete symmetry $g \in G$ which maps $\tilde{x}(t) \in \widetilde{\mathcal{M}}$ to $x(t) \in \mathcal{M}$ .

[chapter 19]

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 (15.11)

$$s(\beta) = \lim_{t\to\infty} \frac{1}{t} \log\langle e^{\beta\cdot(\hat{x}(t)-x)}\rangle_{\mathcal{M}} \,, \tag{24.3}$$

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

$$\frac{\partial}{\partial\beta_i}\frac{\partial}{\partial\beta_j}s(\beta)\bigg|_{\beta=0} = \lim_{t\to\infty}\frac{1}{t}\langle(\hat{x}(t)-x)_i(\hat{x}(t)-x)_j\rangle_{\mathcal{M}} ,$$

yield a (generally anisotropic) diffusion matrix. The spatial diffusion constant is then given by the Einstein relation (15.13)

$$D = \frac{1}{2d}\sum_i \frac{\partial^2}{\partial\beta_i^2}s(\beta)\bigg|_{\beta=0} = \lim_{t\to\infty}\frac{1}{2dt}\langle(\hat{q}(t)-q)^2\rangle_{\mathcal{M}} ,$$

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 to the number of the degrees of freedom.

We now turn to the connection between (24.3) and periodic orbits in the elementary cell. As the full $\hat{\mathcal{M}} \to \widetilde{\mathcal{M}}$ reduction is complicated by the nonabelian nature of $G$, we shall introduce the main ideas in the abelian $\hat{\mathcal{M}} \to \mathcal{M}$ context. [remark 24.6]

### 24.1.1 Reduction from $\hat{\mathcal{M}}$ to $\mathcal{M}$

The key idea follows from inspection of the relation

$$\left\langle e^{\beta\cdot(\hat{x}(t)-x)}\right\rangle_{\mathcal{M}} = \frac{1}{|\mathcal{M}|}\int_{\substack{x\in\mathcal{M}\\ \hat{y}\in\hat{\mathcal{M}}}} dx d\hat{y}\, e^{\beta\cdot(\hat{y}-x)}\delta(\hat{y}-\hat{f}^t(x)) .$$

$|\mathcal{M}| = \int_{\mathcal{M}} dx$ is the volume of the elementary cell $\mathcal{M}$. As in sect. 15.2, 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 $y \in \mathcal{M}$, and $f^t(x)$ given by (24.2). The difference is a translation by a constant, 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:

$$\langle e^{\beta\cdot(\hat{x}(t)-x)}\rangle_{\mathcal{M}} = \frac{1}{|\mathcal{M}|}\int_{x,y\in\mathcal{M}} dx dy\, e^{\beta\cdot(\hat{f}^t(x)-x)}\delta(y-f^t(x)) . \tag{24.4}$$

As this is a translation, the Jacobian is $\delta\hat{y}/\delta y = 1$. In this way the global $\hat{f}^t(x)$ flow averages can be computed by following the flow $f^t(x_0)$ restricted to the elementary cell $\mathcal{M}$. The equation (24.4) suggests that we study the evolution operator

$$\mathcal{L}^t(y, x) = e^{\beta\cdot(\hat{x}(t)-x)}\delta(y-f^t(x)) , \tag{24.5}$$

where $\hat{x}(t) = \hat{f}^t(x) \in \hat{\mathcal{M}}$, but $x$, $f^t(x)$, $y \in \mathcal{M}$. It is straightforward to check that this operator satisfies the semigroup property (15.25),

$$\int_{\mathcal{M}} dz\, \mathcal{L}^{t_2}(y,z)\mathcal{L}^{t_1}(z,x) = \mathcal{L}^{t_2+t_1}(y,x) \ .$$

For $\beta = 0$, the operator (24.5) is the Perron-Frobenius operator (14.10), with the leading eigenvalue $e^{s_0} = 1$ because there is no escape from this system (this will lead to the flow conservation sum rule (20.11) later on).

The rest is old hat. The spectrum of $\mathcal{L}$ is evaluated by taking the trace [section 16.2]

$$\mathrm{tr}\, \mathcal{L}^t = \int_{\mathcal{M}} dx\, e^{\beta\cdot\hat{n}_t(x)}\delta(x - x(t)) \ .$$

Here $\hat{n}_t(x)$ is the discrete lattice translation defined in (24.2). 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, $\hat{f}^{T_p}(x) = x + \hat{n}_p$. In the theory of area–preserving maps such 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 (17.6)

$$\det\left(s(\beta) - \mathcal{A}\right) = \prod_p \exp\left(-\sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{(\beta\cdot\hat{n}_p - sT_p)r}}{\left|\det\left(\mathbf{1} - M_p^r\right)\right|}\right) , \tag{24.6}$$

or the corresponding dynamical zeta function (17.15)

$$1/\zeta(\beta, s) = \prod_p \left(1 - \frac{e^{(\beta\cdot\hat{n}_p - sT_p)}}{|\Lambda_p|}\right) . \tag{24.7}$$

The dynamical zeta function cycle averaging formula (18.21) for the diffusion constant (15.13), zero mean drift $\langle \hat{x}_i \rangle = 0$ , is given by

$$D = \frac{1}{2d} \frac{\left\langle \hat{x}^2 \right\rangle_\zeta}{\langle T \rangle_\zeta} = \frac{1}{2d} \frac{1}{\langle T \rangle_\zeta} \sum{}' \frac{(-1)^{k+1}(\hat{n}_{p_1} + \cdots + \hat{n}_{p_k})^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|} . \tag{24.8}$$

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.

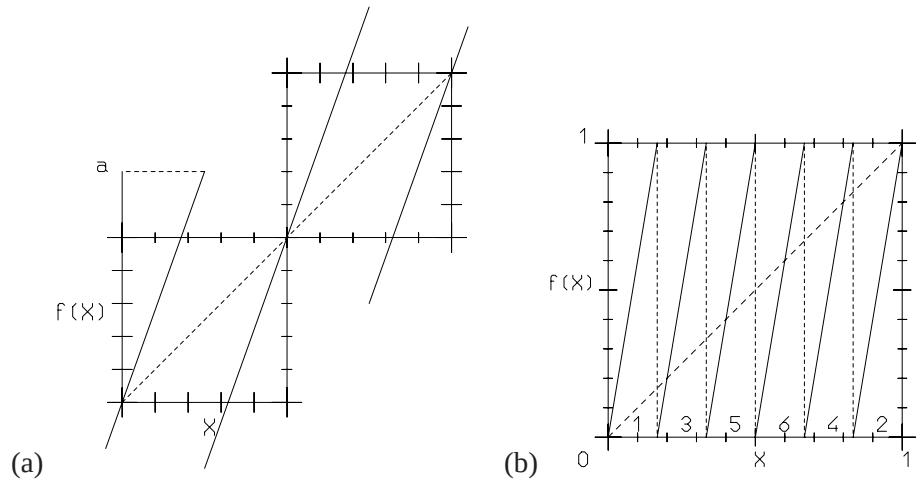



**Figure 24.3:** (a) $\hat{f}(\hat{x})$, the full space sawtooth map (24.9), $\Lambda > 2$. (b) $f(x)$, the sawtooth map restricted to the unit circle (24.12), $\Lambda = 6$.

A sleepy reader might protest that $x_p = x(T_p) - x(0)$ is manifestly equal to zero for a periodic orbit. That is correct; $\hat{n}_p$ in the above formula refers to a displacement on the *infinite* periodic lattice, while $p$ refers to closed orbit of the dynamics reduced to the elementary cell, with $x_p$ belonging to 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 $\left\langle \hat{x}^2 \right\rangle_\zeta$ gets contributions only from the periodic runaway trajectories; they are closed in the elementary cell, but on the periodic lattice each one grows like $\hat{x}(t)^2 = (\hat{n}_p/T_p)^2 = v_p^2 t^2$. 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 $\lambda$ 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-$d$ dynamical systems where all transport coefficients can be evaluated analytically.

## 24.2 Diffusion induced by chains of 1-$d$ 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 10, 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-$d$ expanding maps. This observation motivates introduction of a class of particularly simple 1-$d$ systems, chains of piecewise linear maps.

We start by defining the map $\hat{f}$ on the unit interval as

$$\hat{f}(\hat{x}) = \begin{cases} \Lambda\hat{x} & \hat{x} \in [0, 1/2) \\ \Lambda\hat{x} + 1 - \Lambda & \hat{x} \in (1/2, 1] \end{cases}, \qquad \Lambda > 2\,, \tag{24.9}$$

and then extending the dynamics to the entire real line, by imposing the translation property

$$\hat{f}(\hat{x} + \hat{n}) = \hat{f}(\hat{x}) + \hat{n} \qquad \hat{n} \in \mathbb{Z}\,. \tag{24.10}$$

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

$$\hat{f}(\hat{x}) = -\hat{f}(-\hat{x})\,, \tag{24.11}$$

so the dynamics will exhibit no mean drift; all odd derivatives of the generating function (15.11) with respect to $\beta$, evaluated at $\beta = 0$, will vanish.

The map (24.9) is sketched in figure 24.3 (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 $\Lambda$, 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 (24.10) 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

$$f(x) = \hat{f}(\hat{x}) - \left[\hat{f}(\hat{x})\right], \qquad x = \hat{x} - [\hat{x}] \in [0, 1] \tag{24.12}$$

figure 24.3 (b), where $[\cdots]$ stands for the integer part (24.2). As noted above, 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

$$t_p = z^{n_p} e^{\beta\hat{n}_p}/|\Lambda_p|\,. \tag{24.13}$$

For the piecewise linear map of figure 24.3 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 $\Lambda$ is, the stronger is the stretching action of the map.

The diffusion constant formula (24.8) for 1-$d$ maps is

$$D = \frac{1}{2}\frac{\left\langle \hat{n}^2 \right\rangle_\zeta}{\langle n \rangle_\zeta} \tag{24.14}$$

where the "mean cycle time" is given by (18.22)

$$\langle n \rangle_\zeta = z\frac{\partial}{\partial z}\frac{1}{\zeta(0,z)}\bigg|_{z=1} = -\sum{}' (-1)^k \frac{n_{p_1} + \cdots + n_{p_k}}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}\,, \tag{24.15}$$

and the "mean cycle displacement squared" by (18.25)

$$\left\langle \hat{n}^2 \right\rangle_\zeta = \frac{\partial^2}{\partial \beta^2}\frac{1}{\zeta(\beta,1)}\bigg|_{\beta=0} = -\sum{}' (-1)^k \frac{(\hat{n}_{p_1} + \cdots + \hat{n}_{p_k})^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}\,, \tag{24.16}$$

the primed sum indicating all distinct non-repeating combinations of prime cycles. The evaluation of these formulas in this simple system will require nothing more than pencil and paper.

### 24.2.1 Case of unrestricted symbolic dynamics

Whenever $\Lambda$ is an integer number, the symbolic dynamics is exceedingly simple. For example, for the case $\Lambda = 6$ illustrated in figure 24.3 (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 $\Lambda$ 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 $\{0_+, 1_+, \ldots, (a-1)_+, (a-1)_-, \ldots, 1_-, 0_-\}$, 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 24.3 there are 6 distinct weights (24.13):

$$\begin{aligned} t_1 &= t_2 = z/\Lambda \\ t_3 &= e^{\beta} z/\Lambda\,, \quad t_4 = e^{-\beta} z/\Lambda\,, \quad t_5 = e^{2\beta} z/\Lambda\,, \quad t_6 = e^{-2\beta} z/\Lambda\,. \end{aligned}$$

The piecewise linearity and the simple symbolic dynamics lead to the full cancellation of all curvature corrections in (18.7). The *exact* dynamical zeta function

(13.13) is given by the fixed point contributions:

$$
\begin{aligned}
1/\zeta(\beta,z) &= 1 - t_{0_+} - t_{0_-} - \cdots - t_{(a-1)_+} - t_{(a-1)_-} \\
&= 1 - \frac{z}{a}\left(1 + \sum_{j=1}^{a-1} \cosh(\beta j)\right) . \qquad (24.17)
\end{aligned}
$$

The leading (and only) eigenvalue of the evolution operator (24.5) is

$$
s(\beta) = \log\left\{\frac{1}{a}\left(1 + \sum_{j=1}^{a-1} \cosh(\beta j)\right)\right\} , \qquad \Lambda = 2a,\ a \text{ integer} . \qquad (24.18)
$$

The flow conservation (20.11) 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 (24.14):

$$
\langle n \rangle_\zeta = 2a\frac{1}{\Lambda} = 1 , \qquad \left\langle \hat{x}^2 \right\rangle_\zeta = 2\frac{0^2}{\Lambda} + 2\frac{1^2}{\Lambda} + 2\frac{2^2}{\Lambda} + \cdots + 2\frac{(a-1)^2}{\Lambda} \qquad (24.19)
$$

Using the identity $\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$ we obtain

$$
D = \frac{1}{24}(\Lambda - 1)(\Lambda - 2) , \qquad \Lambda \text{ even integer} . \qquad (24.20)
$$

Similar calculation for odd integer $\Lambda = 2k - 1$ yields [exercise 24.1]

$$
D = \frac{1}{24}(\Lambda^2 - 1) , \qquad \Lambda \text{ odd integer} . \qquad (24.21)
$$

### 24.2.2 Higher order transport coefficients

The same approach yields higher order transport coefficients

$$
\mathcal{B}_k = \frac{1}{k!}\frac{d^k}{d\beta^k} s(\beta)\bigg|_{\beta=0} , \qquad \mathcal{B}_2 = D , \qquad (24.22)
$$

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 generated by the diffusive process. Here $\hat{x}_t$ is the relevant dynamical variable and $\mathcal{B}_k$'s are related to moments $\left\langle \hat{x}_t^k \right\rangle$ of arbitrary order.

Were the diffusive process purely Gaussian

$$
e^{ts(\beta)} = \frac{1}{\sqrt{4\pi Dt}} \int_{-\infty}^{+\infty} d\hat{x}\, e^{\beta\hat{x}} e^{-\hat{x}^2/(4Dt)} = e^{\beta^2 Dt} \qquad (24.23)
$$

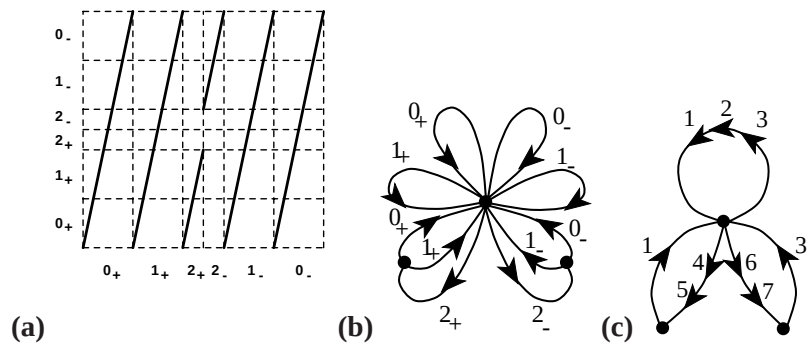



**Figure 24.4:** (a) A partition of the unit interval into six intervals, labeled by the jumping number $\hat{n}(x)$ $I = \{0_+, 1_+, 2_+, 2_-, 1_-, 0_-\}$. The partition is Markov, as the critical point is mapped onto the right border of $\mathcal{M}_{1_+}$. (b) The Markov graph for this partition. (c) The Markov graph in the compact notation of (24.26) (introduced by Vadim Moroz).

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.

For the map under consideration the first Burnett coefficient coefficient $\mathcal{B}_4$ is easily evaluated. For example, using (24.18) in the case of even integer slope $\Lambda = 2a$ we obtain

[exercise 24.2]

$$\mathcal{B}_4 = -\frac{1}{4! \cdot 60}(a-1)(2a-1)(4a^2 - 9a + 7)\,. \tag{24.24}$$

We see that deterministic diffusion is <u>not</u> a Gaussian stochastic process. Higher order even coefficients may be calculated along the same lines.

### 24.2.3 Case of 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.

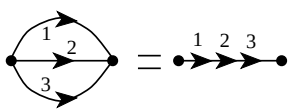
The key idea is to construct a *Markov partition* (10.4), with intervals mapped *onto* unions of intervals. As an example we determine a value of the parameter $4 \le \Lambda \le 6$ for which $f(f(1/2)) = 0$. As in the integer $\Lambda$ case, we partition the unit interval into six intervals, labeled 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 24.4 (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 $f(\mathcal{M}_{1_+}) = \mathcal{M}_{0_+} \bigcup \mathcal{M}_{1_+}$, $f(\mathcal{M}_{2_-}) = \mathcal{M}_{0_-} \bigcup \mathcal{M}_{1_-}$, while the remaining intervals map onto the whole unit interval $\mathcal{M}$. The transition matrix

(10.2) is given by

$$\phi' = T\phi \;=\; \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \phi_{0_+} \\ \phi_{1_+} \\ \phi_{2_+} \\ \phi_{2_-} \\ \phi_{1_-} \\ \phi_{0_-} \end{pmatrix} . \tag{24.25}$$

One could diagonalize (24.25) on a computer, but, as we saw in sect. 10.4, the Markov graph figure 24.4 (b) corresponding to figure 24.4 (a) offers more insight into the dynamics. The graph figure 24.4 (b) can be redrawn more compactly as Markov graph figure 24.4 (c) by replacing parallel lines in a graph by their sum




$$= t_1 + t_2 + t_3 \,. \tag{24.26}$$

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 (13.13) of sect. 13.3, we are led to the dynamical zeta function

$$\begin{aligned} 1/\zeta(\beta, z) &= 1 - t_{0_+} - t_{1_+} - t_{2_+0_+} - t_{2_+1_+} - t_{2_-1_-} - t_{2_-0_-} - t_{1_-} - t_{0_-} \\ &= 1 - \frac{2z}{\Lambda}(1 + \cosh(\beta)) - \frac{2z^2}{\Lambda^2}(\cosh(2\beta) + \cosh(3\beta)) \,. \end{aligned} \tag{24.27}$$

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 (20.11). Conversely, we could have started by picking the desired Markov partition, writing down the corresponding dynamical zeta function, and then fixing $\Lambda$ by the $1/\zeta(0,1) = 0$ condition. For more complicated Markov graphs this approach, together with the factorization (24.35), is helpful in reducing the order of the polynomial condition that fixes $\Lambda$.

The diffusion constant follows from (24.14)

[exercise 24.3]

$$\begin{aligned} \langle n \rangle_\zeta &= 4\frac{1}{\Lambda} + 4\frac{2}{\Lambda^2} \,, \quad \left\langle \hat{n}^2 \right\rangle_\zeta = 2\frac{1^2}{\Lambda} + 2\frac{2^2}{\Lambda^2} + 2\frac{3^2}{\Lambda^2} \\ D &= \frac{15 + 2\sqrt{2}}{16 + 8\sqrt{2}} \,. \end{aligned} \tag{24.28}$$

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 in a finite number of iterations $p$ $f^p(1/2) = 0$. By taking higher and higher values of $p$ 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 (24.9) for a random "generic" value of the parameter $\Lambda$, for example $\Lambda = 4.5$. The idea is to bracket this value of $\Lambda$ 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 $\Lambda = 4.5$. Judging how difficult such problem is already for a tent map (see sect. 13.6), this is not likely to take only a week of work.

Expressions like (24.20) 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 $\Lambda$ of the map (24.9) is increased, or, perhaps more physically, that the diffusion coefficient is a smooth function of the Lyapunov exponent $\lambda$. This turns out not to be true: $D$ as a function of $\Lambda$ is a fractal, nowhere differentiable curve illustrated in figure 24.5. The dependence of $D$ on the map parameter $\Lambda$ is rather unexpected - even though for larger $\Lambda$ more points are mapped outside the unit cell in one iteration, the diffusion constant does not necessarily grow.

This is a consequence of the lack of structural stability, even of purely hyperbolic systems such as the Lozi map and the 1-$d$ diffusion map (24.9). The trouble arises due to non-smooth dependence of the topological entropy on system parameters - any parameter change, no mater 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 $\Lambda$, the number of ways in which the trajectory can re-enter the the initial cell is an irregular function of $\Lambda$.

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 23 have on transport properties of 1-$d$ maps? Consider a $1-d$ map of the real line on itself with the same properties as in sect. 24.2, except for a marginal fixed point at $x = 0$.

A marginal fixed point affects the balance between running and standing orbits, thus generating a mechanism that may result in anomalous diffusion. Our

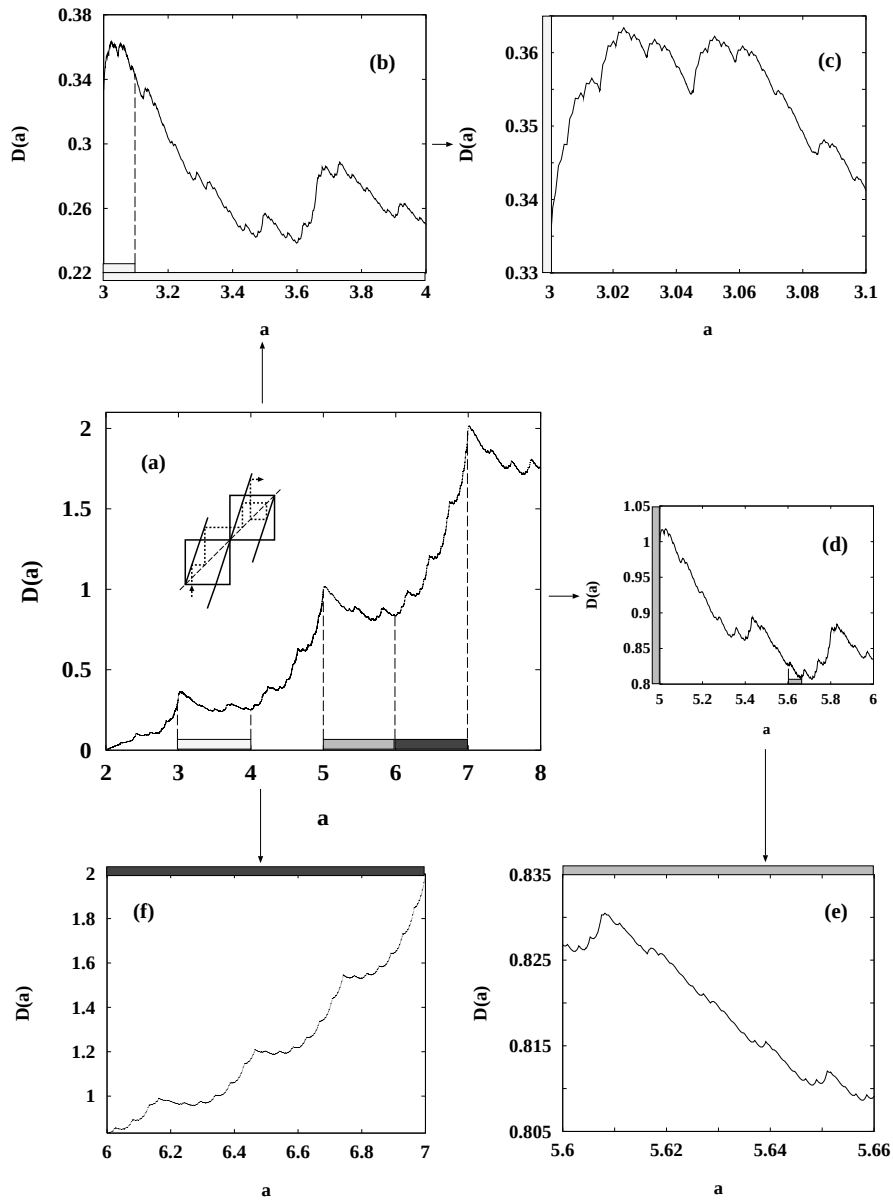



**Figure 24.5:** The dependence of $D$ on the map parameter $a$ is continuous, but not monotone (from ref. [8]). Here $a$ stands for the slope $\Lambda$ in (24.9).

model example is the map shown in figure 24.6 (a), with the corresponding circle map shown in figure 24.6 (b). As in sect. 23.2.1, a branch with support in $\mathcal{M}_i$, $i = 1, 2, 3, 4$ has constant slope $\Lambda_i$, while $f|_{\mathcal{M}_0}$ is of intermittent form. To keep you nimble, this time we take a slightly different choice of slopes. The toy example of sect. 23.2.1 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

[remark 24.8]

$$J(z, s) = \sum_{k=1}^{\infty} z^k/k^s \tag{24.29}$$

Once the $\overline{0}$ fixed point is pruned away, the symbolic dynamics is given by the infinite alphabet $\{1, 2, 3, 4, 0^i1, 0^j2, 0^k3, 0^l4\}$, $i, j, k, l = 1, 2, \ldots$ (compare with

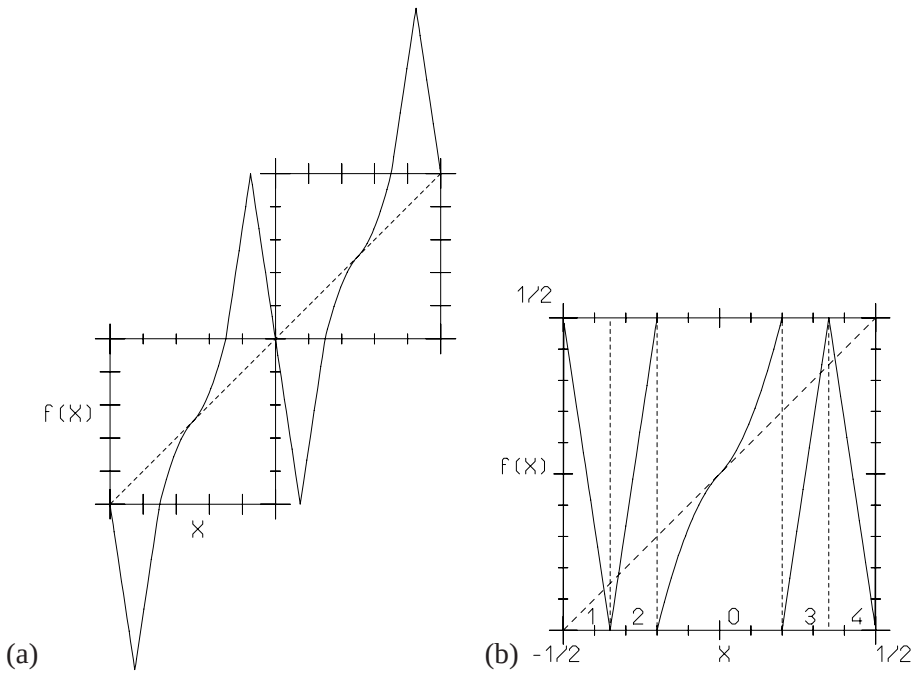



**Figure 24.6:** (a) A map with marginal fixed point. (b) The map restricted to the unit circle.

table **??**). The partitioning of the subinterval $\mathcal{M}_0$ is induced by $\mathcal{M}_{0^k(right)} = \phi^k_{(right)}(\mathcal{M}_3 \bigcup \mathcal{M}_4)$ (where $\phi_{(right)}$ denotes the inverse of the right branch of $\hat{f}|_{\mathcal{M}_0}$) and the same reasoning applies to the leftmost branch. These are regions over which the slope of $\hat{f}|_{\mathcal{M}_0}$ is constant. Thus we have the following stabilities and jumping numbers associated to letters:

$$
\begin{array}{rll}
0^k3,\ 0^k4 & \Lambda_p = \frac{k^{1+\alpha}}{q/2} & \hat{n}_p = 1 \\
0^l1,\ 0^l2 & \Lambda_p = \frac{l^{1+\alpha}}{q/2} & \hat{n}_p = -1 \\
3,\ 4 & \Lambda_p = \pm\Lambda & \hat{n}_p = 1 \\
2,\ 1 & \Lambda_p = \pm\Lambda & \hat{n}_p = -1\,,
\end{array}
\tag{24.30}
$$

where $\alpha = 1/s$ is determined by the intermittency exponent (23.1), while $q$ is to be determined by the flow conservation (20.11) for $\hat{f}$: —PCdefine $R$

$$
\frac{4}{\Lambda} + 2q\zeta(\alpha+1) = 1
$$

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 Jonquiere function (24.29):

$$
1/\zeta_0(z,\beta) \;=\; 1 - \frac{4}{\Lambda} z\cosh\beta - \frac{\Lambda-4}{\Lambda\zeta(1+\alpha)} z\cosh\beta \cdot J(z,\alpha+1)\,. \tag{24.31}
$$

Its first zero $z(\beta)$ is determined by

$$
\frac{4}{\Lambda} z + \frac{\Lambda-4}{\Lambda\zeta(1+\alpha)} z \cdot J(z,\alpha+1) \;=\; \frac{1}{\cosh\beta}\,.
$$

By using implicit function derivation we see that $D$ vanishes (i.e., $\ell'(\beta)|_{\beta=1} = 0$) when $\alpha \leq 1$. The physical interpretation is that a typical orbit will stick for long times near the $\overline{0}$ marginal fixed point, and the 'trapping time' will be larger for higher values of the intermittency parameter $s$ (recall $\alpha = s^{-1}$). Hence, we need 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 (24.7), we may write the diffusion coefficient as an inverse Laplace transform,in such a way that any distinction between maps and flows has vanished. In the case of 1-$d$ diffusion we thus have

$$D = \lim_{t\to\infty} \frac{d^2}{d\beta^2} \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} ds\, e^{st} \frac{\zeta'(\beta, s)}{\zeta(\beta, s)}\bigg|_{\beta=0} \tag{24.32}$$

where the $\zeta'$ 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 $x \to \infty$; 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 $\lim_{t\to\infty} L(ty)/L(t) = 1$. Now

$$\frac{1/\zeta_0{}'(e^{-s},\beta)}{1/\zeta_0(e^{-s},\beta)} = \frac{\left(\frac{4}{\Lambda} + \frac{\Lambda-4}{\Lambda\zeta(1+\alpha)}\left(J(e^{-s},\alpha+1) + J(e^{-s},\alpha)\right)\right)\cosh\beta}{1 - \frac{4}{\Lambda}e^{-s}\cosh\beta - \frac{\Lambda-4}{\Lambda\zeta(1+\alpha)}e^{-s}(e^{-s},\alpha+1)\cosh\beta J}\,.$$

We then take the double derivative with respect to $\beta$ and obtain

$$\frac{d^2}{d\beta^2}\left(1/\zeta_0{}'(e^{-s},\beta)/\zeta^{-1}(e^{-s},\beta)\right)_{\beta=0}$$

$$= \frac{\frac{4}{\Lambda} + \frac{\Lambda-4}{\Lambda\zeta(1+\alpha)}\left(J(e^{-s},\alpha+1) + J(e^{-s},\alpha)\right)}{\left(1 - \frac{4}{\Lambda}e^{-s} - \frac{\Lambda-4}{\Lambda\zeta(1+\alpha)}e^{-s}J(e^{-s},\alpha+1)\right)^2} = g_\alpha(s) \tag{24.33}$$

The asymptotic behavior of the inverse Laplace transform (24.32) 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]

$$\langle (x - x_0)^2 \rangle_t \sim \begin{cases} t & \text{for } \alpha > 1 \\ t^\alpha & \text{for } \alpha \in (0,1) \\ t/\ln t & \text{for } \alpha = 1\,. \end{cases} \tag{24.34}$$

## Résumé

With initial data accuracy $\delta x = |\delta\mathbf{x}(0)|$ and system size $L$, a trajectory is predictable only to the *finite Lyapunov time*

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

Beyond the Lyapunov time chaos rules. Successes of chaos theory: statistical mechanics, quantum mechanics, and questions of long term stability in celestial mechanics.

Tabletop experiment: measure *macroscopic transport* – diffusion, conductance, drag – observe thus determinism on *nanoscales*.

**Chaos: what is it good for?** *TRANSPORT!* Measurable predictions: washboard mean velocity figure 24.7 (a), cold atom lattice figure 24.7 (b), AFM tip drag force figure 24.7 (c).

That Smale's "structural stability" conjecture turned out to be wrong is not a bane of chaotic dynamics - it is actually a virtue, perhaps the most dramatic experimentally measurable prediction of chaotic dynamics. 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.

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

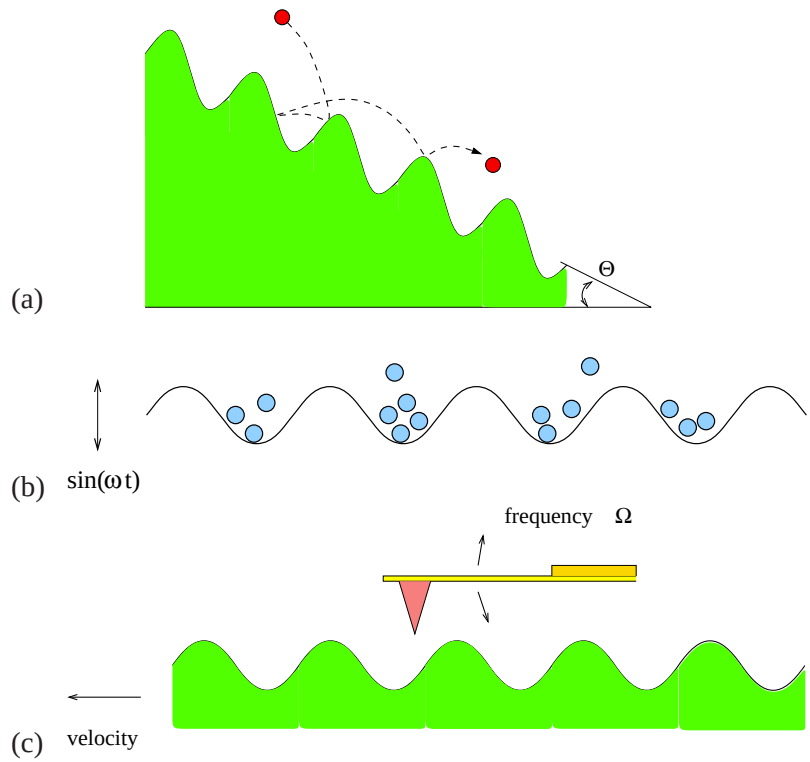



**Figure 24.7:** (a) Washboard mean velocity, (b) cold atom lattice, and (c) AFM tip drag force. (Y. Lan)

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

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

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. 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.

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, and through non-Gaussion relaxation to equilibrium (non-vanishing Burnett coefficients).

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.

Actual evaluation of transport coefficients is a test of the techniques developed above in physical settings. In cases of severe pruning the trace formulas and ergodic sampling of dominant cycles might be more effective strategy than the cycle expansions of dynamical zeta functions and systematic enumeration of all cycles.

## Commentary

**Remark 24.1** Lorentz gas. The original pinball model proposed by Lorentz [4] consisted of randomly, rather than regularly placed scatterers.

**Remark 24.2** Who's dun it? Cycle expansions for the diffusion constant of a particle moving in a periodic array have been introduced independently by R. Artuso [5] (exact dynamical zeta function for 1-$d$ chains of maps (24.8)), by W.N. Vance [6],and by P. Cvitanović, J.-P. Eckmann, and P. Gaspard [7] (the dynamical zeta function cycle expansion (24.8) applied to the Lorentz gas).

**Remark 24.3** Lack of structural stability for D. Expressions like (24.20) 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 $\Lambda$ is a fractal, nowhere differentiable curve shown in figure 24.5. The dependence of $D$ on the map parameter $\Lambda$ is rather unexpected - even though for larger $\Lambda$ more points are mapped outside the unit cell in one iteration, the diffusion constant does not necessarily grow. The fractal dependence of diffusion constant on the map parameter is discussed in refs. [8, 9, 10]. Statistical mechanicians 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. No fractal-like behavior of the conductivity for the Lorentz gas has been detected so far [11].

**Remark 24.4** Diffusion induced by 1-$d$ maps. We refer the reader to refs. [12, 13] for early work on the deterministic diffusion induced by 1-dimenional maps. The sawtooth map (24.9) was introduced by Grossmann and Fujisaka [14] who derived the integer slope formulas (24.20) for the diffusion constant. The sawtooth map is also discussed in refs. [15].

**Remark 24.5** Symmetry factorization in one dimension. In the $\beta = 0$ limit the dynamics (24.11) is symmetric under $x \to -x$, and the zeta functions factorize into products of zeta functions for the symmetric and antisymmetric subspaces, as described in sect. 19.1.1:

$$\frac{1}{\zeta(0,z)} = \frac{1}{\zeta_s(0,z)}\,\frac{1}{\zeta_a(0,z)}\,, \qquad \frac{\partial}{\partial z}\frac{1}{\zeta} = \frac{1}{\zeta_s}\frac{\partial}{\partial z}\frac{1}{\zeta_a} + \frac{1}{\zeta_a}\frac{\partial}{\partial z}\frac{1}{\zeta_s}\,. \tag{24.35}$$

The leading (material flow conserving) eigenvalue $z = 1$ belongs to the symmetric subspace $1/\zeta_s(0,1) = 0$, so the derivatives (24.15) also depend only on the symmetric subspace:

$$\langle n\rangle_\zeta = z\frac{\partial}{\partial z}\frac{1}{\zeta(0,z)}\bigg|_{z=1} = \frac{1}{\zeta_a(0,z)}\,z\frac{\partial}{\partial z}\,\frac{1}{\zeta_s(0,z)}\bigg|_{z=1}\,. \tag{24.36}$$

Implementing the symmetry factorization is convenient, but not essential, at this level of computation.

| length | # cycles | $\zeta(0,0)$ | $\lambda$ |
|---|---|---|---|
| 1 | 5 | -1.216975 | - |
| 2 | 10 | -0.024823 | 1.745407 |
| 3 | 32 | -0.021694 | 1.719617 |
| 4 | 104 | 0.000329 | 1.743494 |
| 5 | 351 | 0.002527 | 1.760581 |
| 6 | 1243 | 0.000034 | 1.756546 |

**Table 24.1:** Fundamental domain, w=0.3 .

**Remark 24.6** Lorentz gas in the fundamental domain. The vector valued nature of the generating function (24.3) 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. [ 16].

In order to perform the full reduction for diffusion one should express the dynamical zeta function (24.7) in terms of the prime cycles of the fundamental domain $\tilde{\mathcal{M}}$ of the lattice (see figure 24.2) 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 $\beta$, or, in other words, the non-commutativity of translations and rotations: see ref. [ 7].

**Remark 24.7** Anomalous diffusion. Anomalous diffusion for 1-$d$ intermittent maps was studied in the continuous time random walk approach in refs. [ 10, 11]. The first approach within the framework of cycle expansions (based on truncated dynamical zeta functions) was proposed in ref. [ 12]. Our treatment follows methods introduced in ref. [ 13], applied there to investigate the behavior of the Lorentz gas with unbounded horizon.

**Remark 24.8** Jonquière functions. In statistical mechanics Jonquière functions

$$J(z,s) = \sum_{k=1}^{\infty} z^k/k^s \tag{24.37}$$

appear in the theory of free Bose-Einstein gas, see refs. [ 22, 23].

## Exercises

24.1. **Diffusion for odd integer $\Lambda$.** Show that when the slope $\Lambda = 2k - 1$ in (24.9) is an odd integer, the diffusion constant is given by $D = (\Lambda^2 - 1)/24$, as stated in (24.21).

24.2. **Fourth-order transport coefficient.** Verify (24.24). 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 (24.28).

24.4. **Maps with variable peak shape:**
Consider the following piecewise linear map

$$f_\delta(x) = \begin{cases} \frac{3x}{1-\delta} & \text{for } x \in \left[0, \frac{1}{3}(1-\delta)\right] \\ \frac{3}{2} - \left(\frac{2}{\delta}\left|\frac{4-\delta}{12} - x\right|\right) & \text{for } x \in \left[\frac{1}{3}(1-\delta), \frac{1}{6}(2+\delta)\right] \\ 1 - \frac{3}{1-\delta}\left(x - \frac{1}{6}(2+\delta)\right) & \text{for } x \in \left[\frac{1}{6}(2+\delta), \frac{1}{2}\right] \end{cases} \tag{24.37}$$

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. [18, 19] 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/pages` offers several project-length deterministic diffusion exercises.

24.6. **Accelerated diffusion.** Consider a map $h$, such that $\hat{h} = \hat{f}$ of figure 24.6 (b), but now running branches are turner 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 $(\phi, \alpha)$ ($\phi$ determines the initial position along the disk, while $\alpha$ 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 $\psi(T)$ 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 $n \mapsto \infty$ $n \simeq T$).

## References


[24.1] J. Machta and R. Zwanzig, *Phys. Rev. Lett.* **50**, 1959 (1983).

[24.2] G.P. Morriss and L. Rondoni, *J. Stat. Phys.* **75**, 553 (1994).

[24.3] L. Rondoni and G.P. Morriss, "Stationary nonequilibrium ensembles for thermostated systems," *Phys. Rev.* **E 53**, 2143 (1996).

[24.4] H.A. Lorentz, *Proc. Amst. Acad.* **7**, 438 (1905).

[24.5] R. Artuso, *Phys. Lett.* **A 160**, 528 (1991).

[24.6] W.N. Vance, *Phys. Rev. Lett.* **96**, 1356 (1992).

[24.7] P. Cvitanović, J.-P. Eckmann, and P. Gaspard, *Chaos, Solitons and Fractals* **6**, 113 (1995).

[24.8] R. Klages, *Deterministic diffusion in one-dimensional chaotic dynamical systems* (Wissenschaft & Technik-Verlag, Berlin, 1996); www.mpipks-dresden.mpg.de/ rklages/publ/phd.html.

[24.9] R. Klages and J.R. Dorfman, *Phys. Rev. Lett.* **74**, 387 (1995); *Phys. Rev. E* **59**, 5361 (1999).

[24.10] R. Klages and J.R. Dorfman, "Dynamical crossover in deterministic diffusion," *Phys. Rev.* **E 55**, R1247 (1997).

[24.11] J. Lloyd, M. Niemeyer, L. Rondoni and G.P. Morriss, *CHAOS* **5**, 536 (1995).

[24.12] T. Geisel and J. Nierwetberg, *Phys. Rev. Lett.* **48**, 7 (1982).

[24.13] M. Schell, S. Fraser and R. Kapral, *Phys. Rev.* **A 26**, 504 (1982).

[24.14] S. Grossmann, H. Fujisaka, Phys. Rev. **A 26**, 1179 (1982); H. Fujisaka and S. Grossmann, Z. Phys. **B 48**, 261 (1982).

[24.15] P. Gaspard and F. Baras, in M. Mareschal and B.L. Holian, eds., *Microscopic simulations of Complex Hydrodynamic Phenomena* (Plenum, NY 1992).

[24.16] F. Christiansen, Master's Thesis, Univ. of Copenhagen (June 1989).

[24.17] P. Cvitanović, P. Gaspard, and T. Schreiber, "Investigation of the Lorentz Gas in terms of periodic orbits," *CHAOS* **2**, 85 (1992).

[24.18] S. Grossmann and S. Thomae, *Phys. Lett.***A 97**, 263 (1983).

[24.19] R. Artuso, G. Casati and R. Lombardi, *Physica* **A 205**, 412 (1994).

[24.20] I. Dana and V.E. Chernov, "Periodic orbits and chaotic-diffusion probability distributions," *Physica A* **332**, 219 (2004).

# Chapter 25

# 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,

$$\rho\left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u}\cdot\nabla\mathbf{u}\right) = -\nabla p + \nu\nabla^2\mathbf{u} + \mathbf{f}\,, \tag{25.1}$$

and solve the problem of turbulence.

Flows described by partial differential equations [PDEs] are said to be 'infinite dimensional' because if one writes them down as a set of ordinary differential equations [ODEs], one needs infinitely many of them to represent the dynamics of one partial differential equation. Even though the state space is infinite-dimensional, the long-time dynamics of many systems of physical interest is finite-dimensional, contained within an *inertial manifold*.

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, $\mathbf{u} \to u$, $\mathbf{u}\cdot\nabla\mathbf{u} \to u\partial_x$, assume constant $\rho$, 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 PDEs, like the one that we turn to now.

## 25.1 Fluttering flame front

Romeo: 'Misshapen chaos of well seeming forms!'
—W. Shakespeare, *Romeo and Julliet*, Act I, Scene I

The Kuramoto-Sivashinsky [KS] system describes the flame front flutter of gas burning on your kitchen stove, figure 25.1 (a), and many other problems of greater import, is one of the simplest nonlinear systems that exhibit 'turbulence' (in this context often referred to more modestly as 'spatiotemporally chaotic behavior'). 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

$$u_t + \tfrac{1}{2}(u^2)_x + u_{xx} + u_{xxxx} = 0\,, \qquad x \in [0,L]\,. \tag{25.2}$$

In this equation $t$ is the time and $x$ is the spatial coordinate. The subscripts $x$ and $t$ denote partial derivatives with respect to $x$ and $t$: $u_t = \partial u/d\partial$, $u_{xxxx}$ 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 (25.1), of the "inertial" term $u\partial_x u$, the "anti-diffusive" term $\partial_x^2 u$ (with a "wrong" sign), etc..

The term $(u^2)_x$ 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 25.1 (b). [section 3.3] [remark 25.1]

Spatial periodicity $u(x,t) = u(x+L,t)$ makes it convenient to work in the Fourier space,

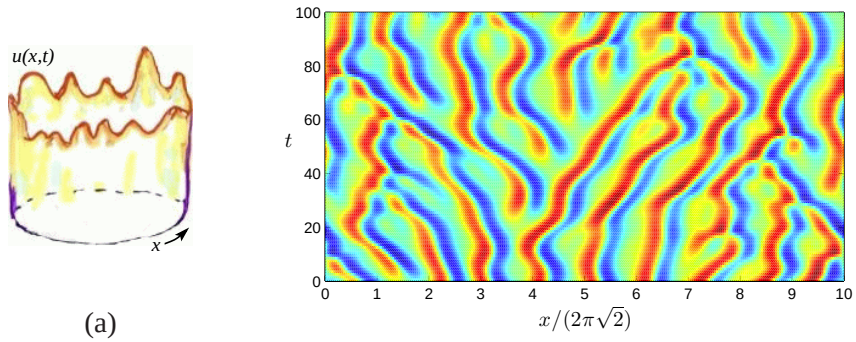
$$u(x,t) = \sum_{k=-\infty}^{+\infty} a_k(t) e^{ikx/\tilde{L}}\,, \tag{25.3}$$

with the 1-dimensional PDE (25.2) replaced by an infinite set of ODEs for the complex Fourier coefficients $a_k(t)$:

$$\dot{a}_k = v_k(a) = ((k/\tilde{L})^2 - (k/\tilde{L})^4)\, a_k - i\frac{k}{2\tilde{L}} \sum_{m=-\infty}^{+\infty} a_m a_{k-m}\,. \tag{25.4}$$

Since $u(x,t)$ is real, $a_k = a^*_{-k}$, and we can replace the sum in (25.10) by a sum over $k > 0$.

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 (25.10)




(a) (b)

**Figure 25.1:** (a) Kuramoto-Sivashinsky dynamics visualized as the Bunsen burner flame flutter, with $u = u(x,t)$ the "velocity of the flame front" at position $x$ and time $t$. (b) A typical "turbulent" solution of the Kuramoto-Sivashinsky equation, system size $L = 88.86$. The color (gray scale) indicates the value of $u$ at a given position and instant in time. The $x$ coordinate is scaled with the most unstable wavelength $2\pi\sqrt{2}$, which is approximately also 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. [10])

to $N$ terms, $16 \le N \le 128$, yield highly accurate solutions for system sizes considered here. Robustness of the Fourier representation of KS as a function of the number of modes kept in truncations of (25.10) is, however, a subtle issue. Adding an extra mode to a truncation of the system 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 $\lambda(\tilde{L}, N)$ for a strange attractor of the system (25.10), there is no reason to expect $\lambda(\tilde{L}, N)$ to smoothly converge to a limit value $\lambda(\tilde{L}, \infty)$ as $N \to \infty$, because of the lack of structural stability both as a function of truncation $N$, and the system size $\tilde{L}$. The topology is more robust for $\tilde{L}$ windows of transient turbulence, where the system can be 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.

Spatial representations of PDEs (such as 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.

The objects explored in this paper: equilibria and short periodic orbits, are robust both under mode truncations and small system parameter $\tilde{L}$ changes.

### 25.1.1 Scaling and symmetries

The Kuramoto-Sivashinsky equation (25.2) is space translationally invariant, time translationally invariant, and invariant under reflection $x \to -x$, $u \to -u$.

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 $v + u(x +$

$2vt, t)$, with $v$ an arbitrary constant velocity, is also a solution. Without loss of generality, in our calculations we shall work in the mean zero velocity frame

$$\int dx\, u = 0\,. \tag{25.5}$$

In terms of the system size $L$, the only length scale available, the dimensions of terms in (25.2) are $[x] = L$, $[t] = L^2$, $[u] = L^{-1}$, $[\nu] = L^2$. Scaling out the "viscosity" $\nu$

$$x \to x\nu^{\frac{1}{2}}\,, \quad t \to t\nu\,, \quad u \to u\nu^{-\frac{1}{2}}\,,$$

brings the Kuramoto-Sivashinsky equation (25.2) to a non-dimensional form

$$u_t = (u^2)_x - u_{xx} - u_{xxxx}\,, \qquad x \in [0, L\nu^{-\frac{1}{2}}] = [0, 2\pi\tilde{L}]\,. \tag{25.6}$$

In this way we trade in the "viscosity" $\nu$ and the system size $L$ for a single dimensionless system size parameter

$$\tilde{L} = L/(2\pi\sqrt{\nu}) \tag{25.7}$$

which plays the role of a "Reynolds number" for the Kuramoto-Sivashinsky system.

In the literature sometimes $L$ is used as the system parameter, with $\nu$ fixed to 1, and at other times $\nu$ is varied with $L$ fixed to either 1 or $2\pi$. To minimize confusion, in what follows we shall state results of all calculations in units of dimensionless system size $\tilde{L}$. Note that the time units also have to be rescaled; for example, if $T_p^*$ is a period of a periodic solution of (25.2) with a given $\nu$ and $L = 2\pi$, then the corresponding solution of the non-dimensionalized (25.6) has period

$$T_p = T_p^*/\nu\,. \tag{25.8}$$

### 25.1.2 Fourier space representation

Spatial periodic boundary condition $u(x,t) = u(x + 2\pi\tilde{L}, t)$ makes it convenient to work in the Fourier space,

$$u(x,t) = \sum_{k=-\infty}^{+\infty} b_k(t) e^{ikx/\tilde{L}}\,. \tag{25.9}$$

with (25.6) replaced by an infinite tower of ODEs for the Fourier coefficients:

$$\dot{b}_k = (k/\tilde{L})^2 \left(1 - (k/\tilde{L})^2\right) b_k + i(k/\tilde{L}) \sum_{m=-\infty}^{+\infty} b_m b_{k-m}\,. \tag{25.10}$$

This is the infinite set of ordinary differential equations promised in this chapter's introduction.

Since $u(x,t)$ is real, $b_k = b^*_{-k}$, so we can replace the sum over $m$ in (25.10) by a sum over $m > 0$. As $\dot{b}_0 = 0$, $b_0$ is a conserved quantity, in our calculations fixed to $b_0 = 0$ by the vanishing mean $\langle u \rangle$ condition (25.5) for the front velocity.

**Example 25.1** ***Kuramoto-Sivashinsky antisymmetric subspace:*** *The Fourier coefficients $b_k$ are in general complex numbers. We can isolate the antisymmetric subspace $u(x,t) = -u(-x,t)$ by considering the case of $b_k$ pure imaginary, $b_k = ia_k$, where $a_k = -a_{-k}$ are real, with the evolution equations*

$$\dot{a}_k = (k/\tilde{L})^2 \left(1 - (k/\tilde{L})^2\right) a_k - (k/\tilde{L}) \sum_{m=-\infty}^{+\infty} a_m a_{k-m} \,. \tag{25.11}$$

*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 (25.11) this corresponds to invariance under*

$$a_{2m} \to a_{2m} \,, a_{2m+1} \to -a_{2m+1} \,. \tag{25.12}$$

*The antisymmetric condition amounts to imposing $u(0,t) = 0$ boundary condition.*

## 25.2 Infinite-dimensional flows: Numerics

The trivial solution $u(x,t) = 0$ is an equilibrium point of (25.2), 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 25.1 (b) computed? The salient feature of such partial differential equations is a theorem saying that for state space contracting flows, the asymptotic dynamics is describable by a *finite* set of "inertial manifold" ordinary differential equations. How you solve the equation (25.2) 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 (25.2) 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 (25.10), 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 *Gälerkin truncations*, or *Gälerkin projections*. You need to worry about "stiffness" of the equations and the stability of your integrator. For [exercise 2.6]

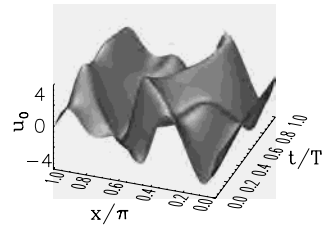



**Figure 25.2:** Spatiotemporally periodic solution $u_0(x,t)$, with period $T_0 = 30.0118$ . The antisymmetric subspace, $u(x,t) = -u(-x,t)$, so we plot $x \in [0, L/2]$. System size $\tilde{L} = 2.89109$, $N = 16$ Fourier modes truncation. (From ref. [4])

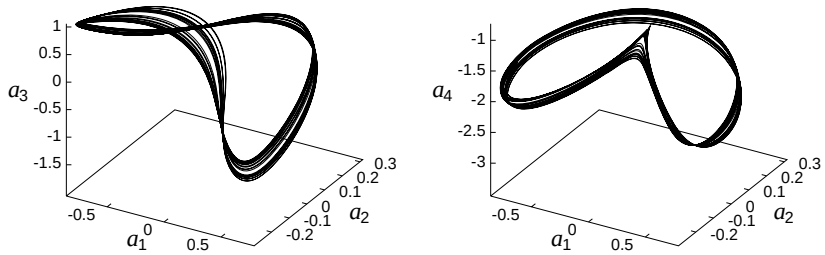



**Figure 25.3:** Projections of a typical 16-dimensional trajectory onto different 3-dimensional subspaces, coordinates (a) $\{a_1, a_2, a_3\}$, (b) $\{a_1, a_2, a_4\}$. System size $\tilde{L} = 2.89109$, $N = 16$ Fourier modes truncation. (From ref. [4].)

the parameter values explored in this chapter, truncations $N$ in range 16 to 64 yields sufficient accuracy.

**Pseudo-spectral methods:** You can mix the two methods, exploiting the speed of Fast Fourier Transforms.

**Example 25.2** ***Kuramoto-Sivashinsky simulation, antisymmetric subspace:*** *To get started, we set $\nu = 0.029910$, $L = 2\pi$ in the Kuramoto-Sivashinsky equation (25.2), or, equivalently, $\nu = 1$, $L = 36.33052$ in the non-dimensionalized (25.6). Consider the antisymmetric subspace (25.11), so the non-dimensionalized system size is $\tilde{L} = L/4\pi = 2.89109$. Truncate (25.11) to $0 \le k \le 16$, 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 (25.9), as in figure 25.1 (b) and figure 25.2. 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.*

## 25.3 Visualization

The problem with high-dimensional representations, such as truncations of the infinite tower of equations (25.10), is that the dynamics is difficult to visualize. The best we can do without much programming is to examine the trajectory's [section 25.3]

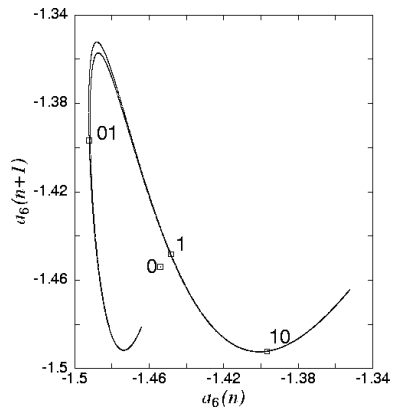



**Figure 25.4:** The attractor of the Kuramoto-Sivashinsky system (25.10), plotted as the $a_6$ component of the $a_1 = 0$ Poincaré section return map. Here 10,000 Poincaré section returns of a typical trajectory are plotted. Also indicated are the periodic points 0, 1, 01 and 10. System size $\tilde{L} = 2.89109$, $N = 16$ Fourier modes truncation. (From ref. [4].)

projections onto any three axes $a_i, a_j, a_k$, as in figure 25.3.

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.

**Example 25.3** ***Kuramoto-Sivashinsky Poincaré 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 (25.10) with $a_1 = 0$, and an arbitrary initial point $(a_2, \ldots, a_N)$. When the flow crosses the $a_1 = 0$ hyperplane in the same direction as initially, the initial point is mapped into $(a'_2, \ldots a'_N) = P(a_2, \ldots, a_N)$. This defines $P$, the Poincaré return map (3.1) of the $(N-1)$-dimensional $a_1 = 0$ hyperplane into itself.*

*Figure 25.4 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.*

The above example 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 25.3, the Poincaré return map figure 25.4 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.

## 25.4 Equilibria of equilibria

(Y. Lan and P. Cvitanović)

The set of equilibria and their stable / unstable manifolds form the coarsest topological framework for organizing state space orbits.

The equilibrium condition $u_t = 0$ for the Kuramoto-Sivashinsky equation PDE (25.6) is the ODE

$$(u^2)_x - u_{xx} - u_{xxxx} = 0$$

which can be analyzed as a dynamical system in its own right. Integrating once we get

$$u^2 - u_x - u_{xxx} = c\,, \tag{25.13}$$

where $c$ is an integration constant whose value strongly influences the nature of the solutions. Written as a 3–$d$ dynamical system with spatial coordinate $x$ playing the role of "time," this is a volume preserving flow

$$u_x = v\,, \qquad v_x = w\,, \qquad w_x = u^2 - v - c\,, \tag{25.14}$$

with the "time" reversal symmetry,

$$x \to -x, \quad u \to -u, \quad v \to v, \quad w \to -w\,.$$

From (25.14) we see that

$$(u + w)_x = u^2 - c\,.$$

If $c < 0$, $u + w$ increases without bound with $x \to \infty$, and every solution escapes to infinity. If $c = 0$, the origin $(0, 0, 0)$ is the only bounded solution.

For $c > 0$ there is much $c$-dependent interesting dynamics, with complicated fractal sets of bounded solutions. The sets of the solutions of the equilibrium condition (25.14) are themselves in turn organized by the equilibria of the equilibrium condition, and the connections between them. For $c > 0$ the equilibrium points of (25.14) are $c_+ = (\sqrt{c}, 0, 0)$ and $c_- = (-\sqrt{c}, 0, 0)$. Linearization of the flow around $c_+$ yields stability eigenvalues $[2\lambda\,, -\lambda \pm i\theta]$ with

$$\lambda = \frac{1}{\sqrt{3}} \sinh\phi\,, \qquad \theta = \cosh\phi\,,$$

and $\phi$ fixed by $\sinh 3\phi = 3\sqrt{3c}$. Hence $c_+$ has a 1–$d$ unstable manifold and a 2–$d$ stable manifold along which solutions spiral in. By the $x \to -x$ "time reversal" symmetry, the invariant manifolds of $c_-$ have reversed stability properties.

The non–wandering set fo 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

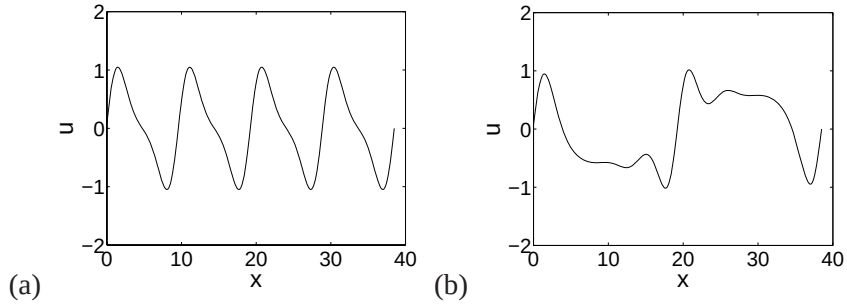



**Figure 25.5:** The non–wandering set under study appears to consist of three patches: the left part ($S_L$), the center part ($S_C$) and the right part ($S_R$), each centered around an unstable equilibrium: (a) central $C_1$ equilibrium, (b) side $R_1$ equilibrium on the interval $[0, L]$.

size $L$ with periodic boundary condition, the only surviving equilibria are those with periodicity $L$. They satisfy the equilibrium condition for (25.10)

$$(k/\tilde{L})^2 \left(1 - (k/\tilde{L})^2\right) b_k + i(k/\tilde{L}) \sum_{m=-\infty}^{+\infty} b_m b_{k-m} = 0 \,. \tag{25.15}$$

Periods of spatially periodic equilibria are multiples of $L$. Every time $\tilde{L}$ crosses an integer value $\tilde{L} = n$, $n$-cell states are generated through pitchfork bifurcations. In the full state space they form an invariant circle due to the translational invariance of (25.6). In the antisymmetric subspace considered here, they corresponds to two points, half-period translates of each other of the form

$$u(x,t) = -2 \sum_k b_{kn} \sin(knx) \,,$$

where $b_{kn} \in \mathbb{R}$.

For any fixed 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 (25.15).

**Example 25.4** ***Some Kuramoto-Sivashinsky equilibria:***

## 25.5 Why does a flame front flutter?

We start by considering the case where $a_q$ is an equilibrium point (2.8). Expanding around the equilibrium point $a_q$, and using the fact that the matrix $\mathbf{A} = \mathbf{A}(a_q)$ in

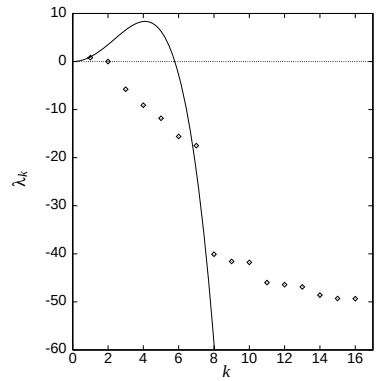



**Figure 25.6:** Lyapunov exponents $\lambda_{\overline{1},k}$ versus $k$ for the least unstable spatio-temporally periodic orbit $\overline{1}$ of the Kuramoto-Sivashinsky system, compared with the Floquet exponents of the $u(x,t) = 0$ stationary solution, $\lambda_k = k^2 - \nu k^4$. The eigenvalue $\lambda_{\overline{1},k}$ for $k \geq 8$ falls below the numerical accuracy of integration and are not meaningful. The cycle $\overline{1}$ was computed using methods of chapter 12. System size $\tilde{L} = 2.89109$, $N = 16$ Fourier modes truncation. (From ref. [4])

(4.2) is constant, we can apply the simple formula (4.30) also to the fundamental matrix of an equilibrium point of a PDE,

$$J^t(a_q) = e^{\mathbf{A}t} \qquad \mathbf{A} = \mathbf{A}(a_q)\,.$$

**Example 25.5** ***Stability matrix, antisymmetric subspace:*** *The Kuramoto-Sivashinsky flat flame front $u(x,t) = 0$ is an equilibrium point of (25.2). The stability matrix (4.3) follows from (25.10)*

$$A_{kj}(a) = \frac{\partial v_k(a)}{\partial a_j} = ((k/\tilde{L})^2 - (k/\tilde{L})^4)\delta_{kj} - 2(k/\tilde{L})a_{k-j}\,. \tag{25.16}$$

*For the $u(x,t) = 0$ equilibrium solution the stability matrix is diagonal, and – as in (4.16) – so is the fundamental matrix $J^t_{kj}(0) = \delta_{kj} e^{((k/\tilde{L})^2 - (k/\tilde{L})^4)t}$.*

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 on which we shall not dwell here.

The $|k| <$?? long wavelength perturbations of the flat-front equilibrium are linearly unstable, while all $|k| >$?? 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. To illustrate the rapid contraction in the non-leading eigendirections we plot in figure 25.6 the eigenvalues of the equilibrium in the unstable regime, for relatively small system size, and compare them with the stability eigenvalues of the least unstable cycle for the same system size. The equilibrium solution is very unstable, in 5 eigendirections, the least unstable cycle only in one. Note that for $k > 7$ the rate of contraction is so strong that higher eigendirections 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 settle converge to the same chaotic attractor.

From (25.10) we see that the origin $u(x,t) = 0$ has Fourier modes as the linear stability eigenvectors. When $|k| \in (0,\tilde{L})$, the corresponding Fourier modes are unstable. The most unstable modes 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, as shown in sect. **??**.

Consider now the case of initial $a_k$ sufficiently small that the bilinear $a_m a_{k-m}$ terms in (25.10) 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 $k^2 > \nu k^4$ is growing with time, and infinitely many with $\nu k^4 > k^2$ 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 in (25.10). 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 (25.10).

Hence, while one may truncate the high modes in the expansion (25.10), 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 (25.2) 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 37 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 (2.2), 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 eigendirections 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 eigendirections.

**Table 25.1:** Important Kuramoto-Sivashinsky equilibria: the first few Floquet exponents

| $S$ | $\mu^{(1)} \pm i\,\omega^{(1)}$ | $\mu^{(2)} \pm i\,\omega^{(2)}$ | $\mu^{(3)} \pm i\,\omega^{(3)}$ |
|---|---|---|---|
| $C_1$ | $0.04422 \pm i\,0.26160$ | $-0.255 \pm i\,0.431$ | $-0.347 \pm i\,0.463$ |
| $R_1$ | $0.01135 \pm i\,0.79651$ | $-0.215 \pm i\,0.549$ | $-0.358 \pm i\,0.262$ |
| $T$ | 0.25480 | $-0.07 \pm i\,0.645$ | -0.264 |

**Example 25.6** ***Stability of Kuramoto-Sivashinsky equilibria:***

spiraling out in a plane*, all other directions contracting*

**Stability of "center" equilibrium**

*linearized Floquet exponents:*

$$(\mu^{(1)} \pm i\,\omega^{(1)}, \mu^{(2)} \pm i\,\omega^{(2)}, \cdots) = (0.044 \pm i\,0.262\,,\ -0.255 \pm i\,0.431\,,\ \cdots)$$

*The plane spanned by* $\mu^{(1)} \pm i\,\omega^{(1)}$ *eigenvectors rotates with angular period* $T \approx 2\pi/\omega^{(1)} = 24.02$.

*a trajectory that starts near the* $C_1$ *equilibrium point spirals away per one rotation with multiplier* $\Lambda_{radial} \approx \exp(\mu^{(1)}T) = 2.9$.

*each Poincaré section return, contracted into the stable manifold by factor of* $\Lambda_2 \approx \exp(\mu^{(2)}T) = 0.002$

*The local Poincaré return map is* in practice $1 - dimensional$

## 25.6 Periodic orbits

expanding eigenvalue of the least unstable spatio-temporally periodic orbit $\overline{1}$: $\Lambda_1 = -2.0\ldots$

*very thin Poincaré section*
thickness $\propto$ least contracting eigenvalue $\Lambda_2 = 0.007\ldots$

15-$d \to$ 15-$d$ Poincaré return map projection on the $[a_6 \to a_6]$ Fourier component is not even $1 \to 1$.

## 25.7 Intrinsic parametrization

Both in the Rössler flow of example 3.4, and in the Kuramoto-Sivashinsky system of example 25.3 we have learned that the attractor is very thin, but otherwise the return maps that we found were disquieting – neither figure 3.6 nor figure 25.4 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

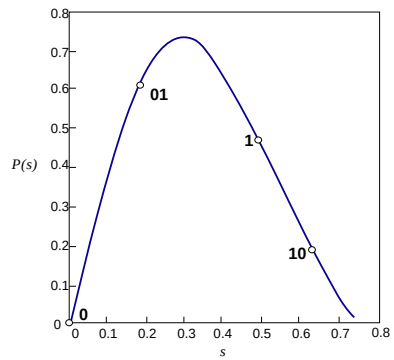



**Figure 25.7:** The Poincaré return map of the Kuramoto-Sivashinsky system (25.10) figure 25.4, from the unstable manifold of the $\overline{1}$ fixed point to the (neighborhood of) the unstable manifold. Also indicated are the periodic points $\overline{0}$ and $\overline{01}$.

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. 12.1), 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 25.3 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 $y_3, y_4 \cdots$. But - thinking is extra price - we do not know how to actually accomplish this.

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 3.6 nor figure 25.4 appeared to be 1-to-1 maps. In this section we show how to use such information to approximately locate cycles.

## 25.8 Energy budget

The space average of a function $a = a(x,t)$ on the interval $L$,

$$\langle a \rangle = \frac{1}{L} \int_0^L dx\, a(x,t)\,, \tag{25.17}$$

is in general time dependent. Its mean value is given by the time average

$$\overline{a} = \lim_{t\to\infty} \frac{1}{t} \int_0^t d\tau\, \langle a \rangle = \lim_{t\to\infty} \frac{1}{tL} \int_0^t \int_0^L d\tau dx\, a(x,\tau)\,. \tag{25.18}$$

The mean value $\overline{a}$, $a = a(u)$ evaluated on an equilibrium or relative equilibrium $u(x,t) = u_q(x - ct)$ is

$$a_q = \langle a \rangle_q \ . \tag{25.19}$$

Evaluation of the infinite time average (25.18) on a function of a period $T_p$ periodic orbit or relative periodic orbit $u_p(x,t)$ requires only a single traversal of the periodic solution,

$$a_p = \frac{1}{T_p} \int_0^{T_p} d\tau \ \langle a \rangle \ . \tag{25.20}$$

Equation (25.2) can be written as

$$u_t = -V_x \,, \qquad V(x,t) = \tfrac{1}{2}u^2 + u_x + u_{xxx} \,. \tag{25.21}$$

$u$ is related to the "flame-front height" $h(x,t)$ by $u = h_x$, so $E$ can be interpreted as the mean energy density (25.22). So, even though KS is a phenomenological small-amplitude equation, the time-dependent quantity

$$E = \frac{1}{L} \int_0^L dx \, V(x,t) = \frac{1}{L} \int_0^L dx \, \frac{u^2}{2} \tag{25.22}$$

has a physical interpretation [**?**] as the average "energy" density of the flame front. This analogy to the corresponding definition of the mean kinetic energy density for the Navier-Stokes will be useful in what follows.

The energy (25.22) is also the quadratic norm in the Fourier space,

$$E = \sum_{k=1}^{\infty} E_k \,, \qquad E_k = \tfrac{1}{2}|a_k|^2 \,. \tag{25.23}$$

Take time derivative of the energy density (25.22), substitute (25.2) and integrate by parts. Total derivatives vanish by the spatial periodicity on the $L$ domain:

$$\begin{aligned} \dot{E} &= \langle u_t \, u \rangle = -\left\langle \left( \frac{u^2}{2} + u \, u_x + u \, u_{xxx} \right)_x u \right\rangle \\ &= \left\langle +u_x \, \frac{u^2}{2} + (u_x)^2 + u_x \, u_{xxx} \right\rangle \,. \end{aligned} \tag{25.24}$$

Substitution by (**??**) verifies that for an equilibrium $E$ is constant:

$$\dot{E} = \left\langle \left( \frac{u^2}{2} + u_x + u_{xxx} \right) u_x \right\rangle = E \, \langle u_x \rangle = 0 \,.$$

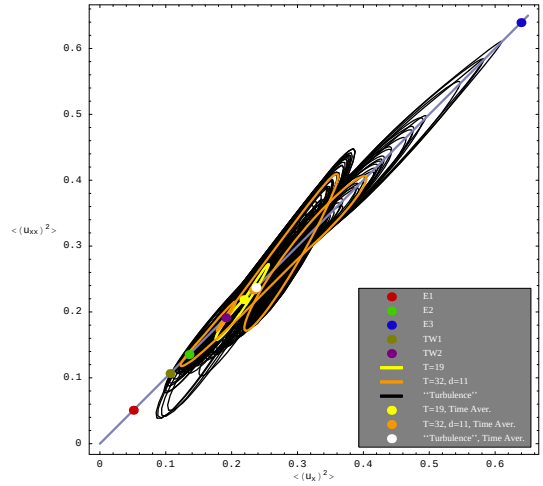



**Figure 25.8:** Power input $\left\langle (u_x)^2 \right\rangle$ vs. dissipation $\left\langle (u_{xx})^2 \right\rangle$ for $L = 22$ equilibria and relative equilibria, for several periodic orbits and relative periodic orbits, and for a typical "turbulent" state. Note that $\overline{(u_{p,x})^2}$ of the $(T_p, d_p) = (32.8, 10.96)$ relative periodic orbit, figure **??**(c), which appears well embedded within the turbulent state, is close to the turbulent expectation $\overline{(u_x)^2}$ .

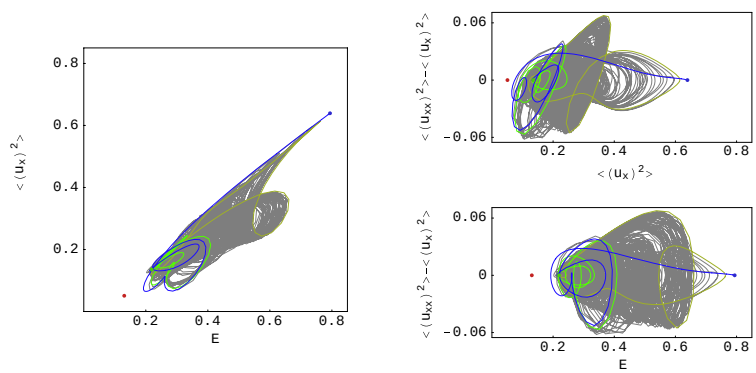



**Figure 25.9:** $E_1$ (red), $E_2$ (green), $E_3$ (blue), connections from $E_1$ to $A(L/4)E_1$ (green), from $A(L/4)E_1$ to $E_1$ (yellow-green) and from $E_3$ to $A(L/4)E_1$ (blue), along with a generic long-time "turbulent" evolution (grey) for $L = 22$. Three different projections of the $(E, \left\langle (u_x)^2 \right\rangle, \left\langle (u_{xx})^2 \right\rangle - \left\langle (u_x)^2 \right\rangle)$ representation are shown.

The first term in (25.24) vanishes by integration by parts, $\left\langle (u^3)_x \right\rangle = 3 \left\langle u_x\, u^2 \right\rangle = 0$ , and integrating the third term by parts yet again we get that the energy variation

$$\dot{E} = \left\langle (u_x)^2 \right\rangle - \left\langle (u_{xx})^2 \right\rangle \tag{25.25}$$

balances the KS equation (25.2) power pumped in by the anti-diffusion $u_{xx}$ against energy dissipated by the hypervicosity $u_{xxxx}$ [**?**].

In figure 25.8 we plot the power input $\left\langle (u_x)^2 \right\rangle$ vs. dissipation $\left\langle (u_{xx})^2 \right\rangle$ for all $L = 22$ equilibria and relative equilibria , several periodic orbits and relative periodic orbits, and for a typical "turbulent" evolution. The time averaged energy density $\overline{E}$ computed on a typical orbit goes to a constant, so the expectation values (25.26) of drive and dissipation exactly balance each out:

$$\overline{\dot{E}} = \lim_{t\to\infty} \frac{1}{t} \int_0^t d\tau\, \dot{E} = \overline{(u_x)^2} - \overline{(u_{xx})^2} = 0\,. \tag{25.26}$$

In particular, the equilibria and relative equilibria sit on the diagonal in figure25.8, and so do time averages computed on periodic orbits and relative periodic orbits:

$$\overline{E}_p \quad = \quad \frac{1}{T_p} \int_0^{T_p} d\tau\, E(\tau)$$

$$\overline{(u_x)^2}_p = \frac{1}{T_p}\int_0^{T_p} d\tau \left\langle (u_x)^2 \right\rangle = \overline{(u_{xx})^2}_p \,. \tag{25.27}$$

In the Fourier basis (25.23) the conservation of energy on average takes form

$$0 = \sum_{k=1}^{+\infty} ((k/\tilde{L})^2 - (k/\tilde{L})^4)\,\overline{E_k}\,, \qquad E_k(t) = |a_k(t)|^2 \,. \tag{25.28}$$

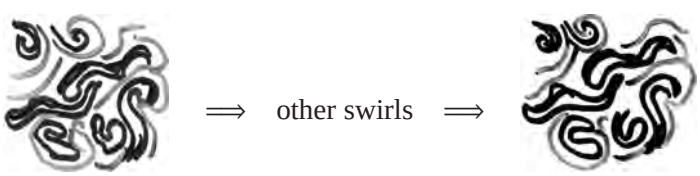
The large $k$ convergence of this series is insensitive to the system size $L$; $\overline{E_k}$ have to decrease much faster than $1/(k/\tilde{L})^4$. Deviation of $E_k$ from this bound for small $k$ determines the active modes. This may be useful to bound the number of equilibria, with the upper bound given by zeros of a small number of long wavelength modes.

## 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 spatially extended systems 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, the system follows approximately for a finite time a 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. Recasting this image into mathematics is the subject of this book.

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. 12.3, are more robust. The relaxation (or [chapter 27]

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 27.

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

**Remark 25.1** Model PDE systems. The theorem on finite dimensionality of inertial manifolds of state space contracting PDE flows is proven in ref. [1]. The Kuramoto-Sivashinsky equation was introduced in refs. [2, 3]. Holmes, Lumley and Berkooz [5] 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. [6] - and this one is perhaps the simplest physically interesting spatially extended nonlinear system.

For equilibria the $L$-independent bound on $E$ is given by Michaelson [**?**]. The best current bound[**?**, **?**] on the long-time limit of $E$ as a function of the system size $L$ scales as $E \propto L^{3/2}$.

The work described in this chapter was initiated by Putkaradze's 1996 term project (see `ChaosBook.org/extras`), and continued by Christiansen Cvitanović, Davidchack, Gibson, Halcrow, Lan, and Siminos [4, 7, 8, 16, 15, 10, 11, 9].

## Exercises

25.1. **Galilean invariance of the Kuramoto-Sivashinsky equation.**

(a) 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, i s also a solution.

(b) Verify that mean

$$\langle u \rangle = \frac{1}{L} \int_L dx\, u$$

is conserved by the flow.

(c) Argue that the choice (25.5) 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?

25.2. **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 $\phi$ 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?

## References


[25.1] C. Foias, B. Nicolaenko, G.R. Sell, and R. Témam, "Kuramoto-Sivashinsky equation," *J. Math. Pures et Appl.* **67**, 197 (1988).

[25.2] Y. Kuramoto and T. Tsuzuki, "Persistent propagation of concentration waves in dissipative media far from thermal equilibrium," *Progr. Theor. Physics* **55**, 365 (1976).

[25.3] G.I. Sivashinsky, "Nonlinear analysis of hydrodynamical instability in laminar flames - I. Derivation of basic equations," *Acta Astr.* **4**, 1177 (1977).

[25.4] F. Christiansen, P. Cvitanović and V. Putkaradze, "Spatiotemporal chaos in terms of unstable recurrent patterns," *Nonlinearity* **10**, 55 (1997); chao-dyn/9606016

[25.5] P. Holmes, J.L. Lumley and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge U. Press, Cambridge 1996).

[25.6] 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. Applied Math.* **50**, 760 (1990).

[25.7] "Chaotic field theory: a sketch," *Physica* **A 288**, 61 (2000) nlin.CD/0001034.

[25.8] Y. Lan, "Dynamical systems approach to 1-*d* spatiotemporal chaos – A cyclist's view," Ph.D. thesis, Georgia Inst. of Tech. (2004).

[25.9] Y. Lan and P. Cvitanović, "Unstable recurrent patterns in Kuramoto-Sivashinsky dynamics," (in preparation, 2007).

[25.10] P. Cvitanović, R. L. Davidchack and E. Siminos, "Topology of a spatio-temporally chaotic Kuramoto-Sivashinsky system" (in preparation, 2007).

[25.11] J. F. Gibson, J. Halcrow, and P. Cvitanović, " On the geometry of state space of a turbulent plane Couette flow: I Exact coherent solutions," (in preparation, 2007).

[25.12] A. K. Kassam and L. N. Trefethen, "Fourth-order time stepping for stiff PDEs," *SIAM J. Sci. Comp.*, (2004).

# Chapter 26

# Noise

He who establishes his argument by noise and command shows that his reason is weak.
—M. de Montaigne

(G. Vattay and P. Cvitanović)

THIS CHAPTER (which reader can safely skip on the first reading) is about noise, how it affects classical dynamics, and the ways it mimics quantum dynamics.

Why - in a study of 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 carries over to weakly stochastic flows in toto. As classical noisy dynamics is more intuitive than quantum dynamics, this exercise 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, not respected by higher order corrections. The same is true of semiclassical quantum dynamics; higher corrections do not respect canonical invariance. Third, the variational principle derived here will be refashioned into a powerful tool for determining periodic orbits in chapter 27.

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, Hamilton-Jacobi formulation, stochastic path integrals.

## 26.1 Deterministic transport

(E.A. Spiegel and P. Cvitanović)

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 this point 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

$$\partial_t I + v \cdot \partial I = 0 \,. \tag{26.1}$$

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 $\mathcal{M}$, $\partial V$ 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,

$$\int_V \left[\partial_t(\rho I) + \partial_i(v_i \rho I)\right] dx = 0 \,,$$

where $\partial_i$ is the partial derivative operator with respect to $x_i$. Since the integration is over an arbitrary volume, we conclude that

$$\partial_t(\rho I) + \partial_i(\rho I v_i) = 0 \,. \tag{26.2}$$

The choice $I \equiv 1$ yields the *continuity equation* for the density:

$$\partial_t \rho + \partial_i(\rho v_i) = 0 \,. \tag{26.3}$$

We have used here the language of fluid mechanics to ease the visualization, but, as we already saw in the discussion of infinitesimal action of the Perron-Frobenius operator (14.25), continuity equation applies to any deterministic state space flow.

## 26.2 Brownian diffusion

Consider tracer molecules, let us say green molecules, embedded in a denser gas of transparent molecules. Assume that the density of tracer molecules $\rho$ compared to the background gas density 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 $j_i$ flow through its surface $\partial V$ (with $\hat{n}$ its outward normal):

$$\partial_t \int_V dx\, \rho = -\int_{\partial V} d\sigma\, \hat{n}_i\, j_i\,. \tag{26.4}$$

The divergence theorem turns this into the conservation law for tracer density:

$$\partial_t \rho + \partial_i j_i = 0\,. \tag{26.5}$$

The tracer density $\rho$ 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 $j$? 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 $\rho v_i$ in (26.3)) is driven by the density gradient

$$j_i = -D\frac{\partial \rho}{\partial x_i}\,. \tag{26.6}$$

This is the *Fick law*, with the diffusion constant $D$ a phenomenological parameter. For simplicity here we assume that $D$ is a scalar; in general $D \to D_{ij}(x,t)$ is a space- and time-dependent tensor. Substituting this $j$ into (26.5) yields the *diffusion equation*

$$\frac{\partial}{\partial t}\rho(x,t) = D\frac{\partial^2}{\partial x^2}\rho(x,t)\,. \tag{26.7}$$

This linear equation has an exact solution in terms of an initial Dirac delta density distribution, $\rho(x,0) = \delta(x - x_0)$,

$$\rho(x,t) = \frac{1}{(4\pi Dt)^{d/2}} e^{-\frac{(x-x_0)^2}{4Dt}} = \frac{1}{(4\pi Dt)^{d/2}} e^{-\frac{x^2}{4Dt}}\,. \tag{26.8}$$

The average distance covered in time $t$ obeys the Einstein diffusion formula

$$\left\langle (x - x_0)^2 \right\rangle_t = \int dx\, \rho(x,t)(x - x_0)^2 = 2dDt\,. \tag{26.9}$$

## 26.3 Weak noise

> 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

So far we have considered tracer molecule dynamics which is purely Brownian, with no deterministic "drift." Consider next a deterministic flow $\dot{x} = v(x)$ perturbed by a stochastic term $\xi(t)$,

$$\dot{x} = v(x) + \xi(t)\,. \tag{26.10}$$

Assume that $\xi(t)$'s fluctuate around $[\dot{x} - v(x)]$ with a Gaussian probability density

$$p(\xi, \delta t) = \left(\frac{\delta t}{4\pi D}\right)^{d/2} e^{-\frac{\xi^2}{4D}\delta t}\,, \tag{26.11}$$

and are uncorrelated in time (white noise)

$$\langle \xi(t)\xi(t')\rangle = 2dD\delta(t-t')\,. \tag{26.12}$$

The normalization factors in (26.8) and (26.11) differ, as $p(\xi, \delta t)$ is a probability density for velocity $\xi$, and $\rho(x,t)$ is a probability density for position $x$. The material particle now drifts along the trajectory $x(t)$, so the velocity diffusion follows (26.8) for infinitesimal time $\delta t$ only. As $D \to 0$, the distribution tends to the (noiseless, deterministic) Dirac delta function.

An example is the Langevin 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.

The phenomenological Fick law current (26.6) is now a sum of two components, the material particle center-of-mass deterministic drift $v(x)$ and the weak noise term

$$j_i = \rho v_i - D\frac{\partial \rho}{\partial x_i}\,, \tag{26.13}$$

Substituting this $j$ into (26.5) yields the *Fokker-Planck equation*

$$\partial_t \rho + \partial_i(\rho v_i) = D\,\partial^2 \rho. \tag{26.14}$$

The left hand side, $d\rho/dt = \partial_t\rho + \partial\cdot(\rho v)$, is deterministic, with the continuity equation (26.3) recovered in the weak noise limit $D \to 0$. 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 $\rho$. Where is the density going globally?

If the system is bound, the probability density vanishes sufficiently fast outside the central region, $\rho(x,t) \to 0$ as $|x| \to \infty$, 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 invariant density

$$\rho_0(x) = \lim_{t\to\infty}\rho(x,t)\,, \tag{26.15}$$

an eigenfunction $\rho(x,t) = e^{st}\rho_0(x)$ of the time-independent Fokker-Planck equation

$$\left(\partial_i v_i - D\,\partial^2 + s_\alpha\right)\rho_\alpha = 0\,, \tag{26.16}$$

with vanishing eigenvalue $s_0 = 0$. Provided the noiseless classical flow is hyperbolic, in the vanishing noise limit the leading eigenfunction of the Fokker-Planck equation tends to natural measure (14.17) 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 (26.16) 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 $\gamma$ for which any initial probability density converges to a finite equilibrium density is called the *escape rate*. In the noiseless limit this coincides with the deterministic escape rate (15.15).

We have introduced noise phenomenologically, and used the weak noise assumption in retaining only the first derivative of $\rho$ in formulating the Fick law (26.6) and including noise additively in (26.13). A full theory of stochastic ODEs is much subtler, but this will do for our purposes.

## 26.4 Weak noise approximation

In the spirit of the WKB approximation, we shall now study the evolution of the probability distribution by rewriting it as

$$\rho(x,t) = e^{\frac{1}{2D}R(x,t)} \,. \tag{26.17}$$

The time evolution of $R$ is given by

$$\partial_t R + v\partial R + (\partial R)^2 = D\partial v + D\partial^2 R \,.$$

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 the Hamilton-Jacobi equation . The function $R$ can be interpreted as the Hamilton's principal function, corresponding to the Hamiltonian

$$H(x,p) = p\,v(x) + p^2/2 \,,$$

with the Hamilton's equations of motion

$$\begin{aligned} \dot{x} &= \partial_p H = v + p \\ \dot{p} &= -\partial_x H = -A^T p \,, \end{aligned} \tag{26.18}$$

where $A$ is the stability matrix (4.3)

$$A_{ij}(x) = \frac{\partial v_i(x)}{\partial x_j} \,.$$

The noise Lagrangian is then

$$L(x,\dot{x}) = \dot{x}\cdot p - H = \frac{1}{2}\,[\dot{x} - v(x)]^2 \,. \tag{26.19}$$

We have come the full circle - the Lagrangian is the exponent of our assumed Gaussian distribution (26.11) 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 (26.11) 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)

$$
\begin{aligned}
P(x, x_0, t) \quad &\sim \quad \sum_{\mathcal{P}} \prod_j p(\xi(\tau_j), \delta\tau_j) = \int \prod_j d\xi_j \left(\frac{\delta\tau_j}{2\pi D}\right)^{d/2} e^{-\frac{\xi(\tau_j)^2}{2D}\delta\tau_i} \\
&\to \quad \frac{1}{Z} \sum_{\mathcal{P}} \exp\left(-\frac{1}{2D}\int_0^t d\tau\, \xi^2(\tau)\right), \qquad (26.20)
\end{aligned}
$$

where $\delta\tau_i = \tau_i - \tau_i$, 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 (26.10) 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 the probability is maximized by a solution of Hamilton's equations (26.18) that connects the two points $x_0 \to x'$ in time $t$.

## 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 $D \to 0$. The effect of the noise can be taken into account by adding noise corrections to the classical trace formula.

## Commentary

**Remark 26.1** Literature. The theory of stochastic processes is a vast subject, spanning over centuries and over disciplines ranging from pure mathematics to impure finance. We enjoyed reading van Kampen classic [1], 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 [ 2] will do just as well. The "Langevin equation" introduces noise and damping only into the acceleration of Newton's equations; here we are considering more general stochastic differential equations in the weak noise limit. Onsager-Machlup seminal paper [ 18] was the first to introduce a variational method - the "principle of least dissipation" - based on the Lagrangian of form (26.19). This paper deals only with a finite set of linearly damped thermodynamic variables. Here the setting is much more general: we study fluctuations over a state space varying velocity field $v(x)$. Schulman's monograph [ 11] contains a very readable summary of Kac's [12] exposition of Wiener's integral over stochastic paths.

## Exercises

26.1. **Who ordered $\sqrt{\pi}$ ?** Derive the Gaussian integral

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx\, e^{-\frac{x^2}{2a}} = \sqrt{a}\,, \qquad a > 0\,.$$

assuming only that you know to integrate the exponential function $e^{-x}$. Hint, hint: $x^2$ is a radius-squared of something. $\pi$ is related to the area or circumference of something.

26.2. ***D*-dimensional Gaussian integrals.** Show that the Gaussian integral in $D$-dimensions is given by

$$\frac{1}{(2\pi)^{d/2}}\int d^d\phi e^{-\frac{1}{2}\phi^T\cdot M^{-1}\cdot\phi+\phi\cdot J} = |\det M|^{\frac{1}{2}}\, e^{\frac{1}{2}J^T\cdot M\cdot J} \tag{26.21}$$

where $M$ is a real positive definite $[d\times d]$ matrix, i.e., a matrix with strictly positive eigenvalues. $x$, $J$ are $D$-dimensional vectors, and $x^T$ is the transpose of $x$.

26.3. **Convolution of Gaussians.** Show that the Fourier transform of convolution

$$[f*g](x) = \int d^d y\, f(x-y)g(y)$$

of two Gaussians

$$f(x) = e^{-\frac{1}{2}x^T\cdot\frac{1}{\Delta_1}\cdot x}\,, \qquad g(x) = e^{-\frac{1}{2}x^T\cdot\frac{1}{\Delta_2}\cdot x}$$

factorizes as

$$[f*g](x) = \frac{1}{(2\pi)^d}\int dk\, F(k)G(k)e^{ik\cdot x}\,, \tag{26.22}$$

where

$$\begin{aligned} F(k) &= \frac{1}{(2\pi)^d}\int d^d x\, f(x)e^{-ik\cdot x} = |\det\Delta_1|^{1/2}e^{\frac{1}{2}k^T\cdot\Delta_1\cdot k} \\ G(k) &= \frac{1}{(2\pi)^d}\int d^d x\, g(x)e^{-ik\cdot x} = |\det\Delta_2|^{1/2}e^{\frac{1}{2}k^T\cdot\Delta_2\cdot k} \end{aligned}$$

Hence

$$\begin{aligned} [f*g](x) &= \frac{1}{(2\pi)^d}|\det\Delta_1\det\Delta_1|^{1/2}\int d^d p\, e^{\frac{1}{2}p^T\cdot(\Delta_1+\Delta} \\ &= \left|\frac{\det\Delta_1\det\Delta_2}{\det(\Delta_1+\Delta_2)}\right|^{1/2} e^{-\frac{1}{2}x^T\cdot(\Delta_1+\Delta_2)^{-1}\cdot x}\,. \end{aligned}$$

## References


[26.1] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North Holland, Amsterdam, 1981).

[26.2] H. Risken, *The Fokker-Planck equation: Methods of solution and applications* (Springer-Verlag, New York, 1989).

[26.3] W. Dittrich and M. Reuter, *Classical and Quantum Dynamics: From Classical Paths to Path Integrals* (Springer-Verlag, Berlin 1994).

[26.4] E. M. Lifshitz and L. P. Pitaevskii, *Physical Kinetics* (Pergamon, London 1981).

[26.5] C. Itzykson and J.-M. Drouffe, *Statistical field theory* (Cambridge U. Press, 1991).

[26.6] V. Ambegaokar, *Reasoning about luck; probability and its uses in physics* (Cambridge Univ. Press, Cambridge 1996).

[26.7] B. Sakita, *Quantum theory of many variable systems and fields* (World Scientific, Singapore 1985).

[26.8] G. E. Uhlenbeck, G. W. Ford and E. W. Montroll, *Lectures in Statistical Mechanics* (Amer. Math. Soc., Providence R.I., 1963).

[26.9] M. Kac, "Random walk and the theory of Brownian motion," (1946), reprinted in ref. [10].

[26.10] N. Wax, ed., *Selected Papers on Noise and Stochastic Processes* (Dover, New York 1954).

[26.11] L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley-Interscience, New York 1981).

[26.12] M. Kac, *Probability and Related Topics in Physical Sciences* (Wiley-Interscience, New York 1959).

[26.13] E. Nelson, *Quantum Fluctuations* (Princeton Univ. Press 1985).

[26.14] H. Kunita, *Stochastic Flows and Stochastic Differential Equations* (Cambridge Univ. Press, 1990).

[26.15] H. Haken and G. Mayer-Kress, *Z. f. Physik* **B 43**, 185 (1981).

[26.16] M. Roncadelli, *Phys. Rev.* **E 52**, 4661 (1995).

[26.17] G. Ryskin, *Phys. Rev.* **E 56**, 5123 (1997).

[26.18] L. Onsager and S. Machlup, *Phys. Rev.* **91**, 1505, 1512 (1953).

[26.19] Lord Rayleigh, *Phil. Mag.* **26**, 776 (1913).

[26.20] E. Gozzi, M. Reuter and W. D. Thacker, *Phys. Rev.* **D 40**, 3363 (1989).

[26.21] E. Gozzi and M. Reuter, *Phys. Lett.* **233B**, 383 (1989); **238B**, 451 (1990); **240B**, 137 (1990).

[26.22] E. Gozzi, M. Reuter and W. D. Thacker, *Phys. Rev.* **D 46**, 757 (1992).

[26.23] E. Gozzi, M. Reuter and W. D. Thacker, *Chaos, Solitons and Fractals* **2**, 441 (1992).

[26.24] Benzi et al., *Jour. Phys.* **A 18**, 2157 (1985).

[26.25] R. Graham, *Europhys. Lett.* **5**, 101 (1988).

[26.26] R. Graham and T. Tél, *Phys. Rev. A* **35**, 1382 (1987).

[26.27] R. Benzi, G. Paladin, G. Parisi and A. Vulpiani, *J. Phys.* **A 18**, 2157 (1985).

[26.28] L. Arnold and V. Wihstutz, *Lyapunov exponents, Lecture Notes in Math.* **1186** (Springer-Verlag, New York 1986).

[26.29] M. V. Feigelman and A. M. Tsvelik, *Sov. Phys. JETP* **56**, 823 (1982).

[26.30] G. Parisi and N. Sourlas, *Nucl. Phys.* **B206**, 321 (1982).

[26.31] F. Langouche et al., *Functional integration and semiclassical expansion* (Reidel, Dordrecht 1982).

[26.32] O. Cepas and J. Kurchan, "Canonically invariant formulation of Langevin and Fokker-Planck equations," *European Phys. J.* **B 2**, 221 (1998), cond-mat/9706296.

[26.33] S. Tanase-Nicola and J. Kurchan, "Statistical-mechanical formulation of Lyapunov exponents," cond-mat/0210380.

# Chapter 27

# Relaxation for cyclists

CYCLES, i.e., solutions of the periodic orbit condition (12.1)

$$f^{t+T}(x) = f^t(x)\,, \qquad T > 0 \tag{27.1}$$

are prerequisite to chapters 16 and 17 evaluation of spectra of classical evolution operators.Chapter 12 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. 12.2.1 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 (12.5)

$$x'_i = g_i(x) = x_i - \left(\frac{1}{M(x) - \mathbf{1}}\right)_{ij} (f(x) - x)_j\,. \tag{27.2}$$

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 12 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 (27.1). 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. 27.1 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 $(\tilde{x}_1^{(0)}, \tilde{x}_2^{(0)}, \cdots, \tilde{x}_N^{(0)})$ 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

$$F^2(\tilde{x}^{(k)}) = \sum_i^N \left(\tilde{x}_{i+1}^{(k)} - f(\tilde{x}_i^{(k)})\right)^2 \tag{27.3}$$

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. 27.1 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. 27.2 we replace such flows by discrete iterations. In sect. 27.3 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.

## 27.1 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),

$$\begin{aligned} x_{i+1} &= 1 - ax_i^2 + by_i \\ y_{i+1} &= x_i \,, \end{aligned} \tag{27.4}$$

is a stationary solution of the relaxation dynamics defined by the flow

$$\frac{dx_i}{d\tau} = v_i, \quad i = 1, \ldots, n \tag{27.5}$$

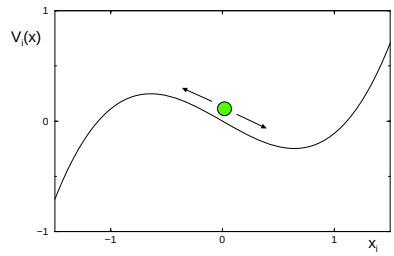



**Figure 27.1:** "Potential" $V_i(x)$ (27.7) for a typical point along an initial guess trajectory. For $\sigma_i = +1$ the flow is toward the local maximum of $V_i(x)$, and for $\sigma_i = -1$ toward the local minimum. A large deviation of $x_i$'s is needed to destabilize a trajectory passing through such local extremum of $V_i(x)$, hence the basin of attraction is expected to be large.

for any vector field $v_i = v_i(x)$ which vanishes on the trajectory. Here $\tau$ 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)

$$v_i = x_{i+1} - 1 + a x_i^2 - b x_{i-1}. \tag{27.6}$$

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 $i$)

$$v_i = \frac{\partial}{\partial x_i} V_i(x)\,, \qquad V_i(x) = x_i(x_{i+1} - b x_{i-1} - 1) + \frac{a}{3} x_i^3\,. \tag{27.7}$$

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 (27.5) by

$$\frac{dx_i}{d\tau} = \sigma_i v_i, \quad i = 1, \ldots, n, \tag{27.8}$$

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 (27.8) - all $x_i$ and $V_i(x)$ change at each integration step - but this is the observation that motivates the method. The differential equations (27.8) 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 27.1. As the "potential" function (27.7) 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 27.1 of order 1, in contrast to the exponential precision required of initial guesses by the Newton-Raphson method.

**Example 27.1** ***Hénon map cycles.*** *Our aim in this calculation is to find all periodic orbits of period $n$ for the Hénon map (27.4), in principle at most $2^n$ orbits. We start by choosing an initial guess trajectory $(x_1, x_2, \cdots, x_n)$ and impose the periodic boundary*

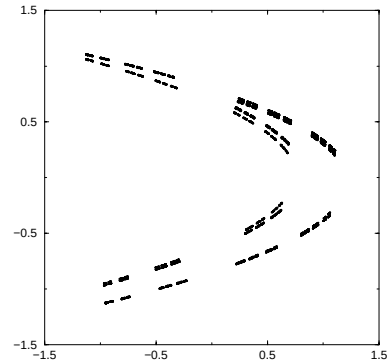



**Figure 27.2:** The repeller for the Hénon map at $a = 1.8$, $b = 0.3$ .

*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* **??**. *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 27.1 gives some indication of a typical basin of attraction of the method (see also figure 27.3).*

*The calculation is carried out by solving the set of $n$ ordinary differential equations (27.8) using a simple Runge-Kutta method with a relatively large step size ($h = 0.1$) until $|v|$ becomes smaller than a given value $\varepsilon$ (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 27.3]

*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* **??**. *The number of unstable periodic orbits for periods $n \leq 28$ is given in table* **??**. *Comparing this with the list of all possible 2-symbol alphabet prime cycles, table* **??**, *we see that the pruning is quite extensive, with the number of cycle 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 27.2. Comparing this repelling set with the strange attractor for the Hénon's parameters figure 3.9, we note the existence of gaps in the set, cut out by the preimages of the escaping regions.*

[remark 27.2]

*In practice, the relaxation flow (27.8) 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 length and accuracy.*

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

$$\frac{dx}{d\tau} = f(x) - x. \tag{27.9}$$

**Table 27.1:** 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 27.4), a cycle 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 $\overline{0}$ cycle 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 cycle 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.

| $n$ | $p$ | $(y_p, x_p)$ | $\lambda_p$ |
|---|---|---|---|
| 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 27.2:** 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 length $n$, 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 |

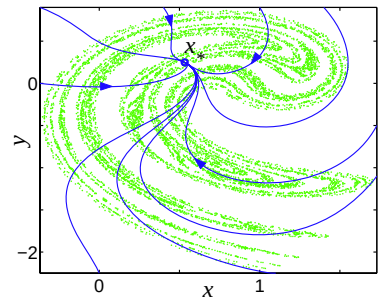



**Figure 27.3:** Typical trajectories of the vector field (27.9) for the stabilization of a hyperbolic fixed point of the Ikeda map (27.11) located at $(x, y) \approx (0.53275, 0.24689)$. The circle indicates the position of the fixed point. Note that the basin of attraction of this fixed point is large, larger than the entire Ikeda attractor.

If all eigenvalues of the fundamental matrix $J(x_*) = Df(x_*)$ have real parts smaller than unity, then $x_*$ is a stable equilibrium point of the flow.

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 (27.8), modify the flow by

$$\frac{dx}{d\tau} = \mathbf{C}\,(f(x) - x)\ , \tag{27.10}$$

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 $\pm 1$ 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.

**Example 27.2** ***Ikeda map:*** *We illustrate the method with the determination of the periodic orbits of the Ikeda map:*

$$\begin{aligned} x' &= 1 + a(x\cos w - y\sin w) \\ y' &= a(x\sin w + y\cos w) \\ \text{where}\ \ w &= b - \frac{c}{1 + x^2 + y^2}\ , \end{aligned} \tag{27.11}$$

*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 fundamental matrix $(\Lambda_1, \Lambda_2) \approx (-2.3897, -0.3389)$, so the flow is already stabilized with $\mathbf{C} = \mathbf{1}$. Figure 27.3 depicts the flow of the vector field around the fixed point $x_*$.*

*In order to determine $x_*$, one needs to integrate the vector field (27.9) 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 $\mathbf{C} = \mathbf{1}$, Figure 27.4 (a), indicates a hyperbolic equilibrium point, while for $\mathbf{C} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ the flow of the vector field, figure 27.4 (b) indicates that $x_*$ is an attracting equilibrium point, reached at exponential speed by integration forward in time.*

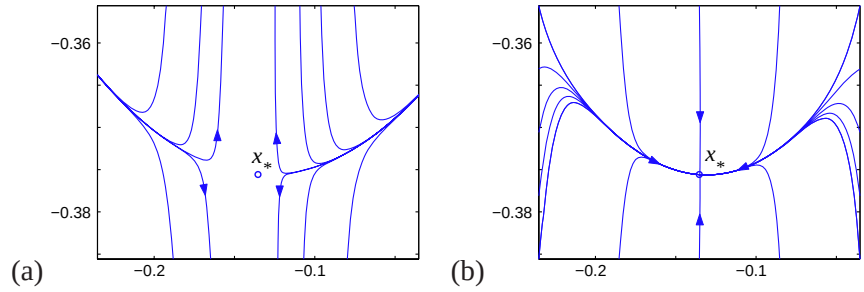



**Figure 27.4:** Typical trajectories of the vector field (27.10) for a hyperbolic fixed point $(x,y) \approx (-0.13529, -0.37559)$ of $f^3$, where $f$ is the Ikeda map (27.11). The circle indicates the position of the fixed point. For the vector field corresponding to (*a*) $\mathbf{C} = \mathbf{1}$, $x_*$ is a hyperbolic equilibrium point of the flow, while for (*b*) $\mathbf{C} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, $x_*$ is an attracting equilibrium point.

The generalization from searches for fixed points to searches for cycles is straightforward. In order to determine a prime cycle $x = (x_1, x_2, \ldots, x_n)$ of a $d$-dimensional map $x' = f(x)$, we modify the multipoint shooting method of sect. 12.3, and consider the $nd$-dimensional vector field

$$\frac{dx}{d\tau} = \mathbf{C}\,(f(x) - x)\ , \tag{27.12}$$

where $f(x) = (f(x_n), f(x_1), f(x_2), \ldots, f(x_{n-1}))$, 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 $\pm 1$. Risking a bit of confusion, we denote by $x$, $f(x)$ both the $d$-dimensional vectors in (27.10), and $nd$-dimensional vectors in (27.12), as the structure of the equations is the same.

## 27.2 Discrete iteration relaxation method

(C. Chandre, F.K. Diakonos and P. Schmelcher)

The problem with the Newton-Raphson iteration (27.2) 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 length $n$ 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 (27.8) 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 (27.9) 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 fundamental matrix prefactor in (27.2) (whose inversion can be time-consuming) by a constant matrix prefactor

$$x' = g(x) = x + \Delta\tau \mathbf{C}(f(x) - x), \tag{27.13}$$

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 $\pm 1$ 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 $\Delta\tau \to d\tau$ the map (27.13) is the Euler method for integrating the modified flow (27.10), 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. 27.1. Indeed, the flow (27.8) is the simplest example of this method, with the infinitesimal fictitious time increment $\Delta\tau \to d\tau$, the infinitesimal coordinate correction $(x - x') \to dx_i$, and the $[n\times n]$ diagonal matrix $\mathbf{C} \to \sigma_i = \pm 1$.

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 matrices [remark 27.3]

$$\mathbf{C} \in \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right\}$$

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 $\alpha$ 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 on the attractor on 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 (27.2).

The fixed points of the $n$th iterate $f^n$ are cycle points of a cycle of period $n$. If we consider the map

$$x' = g(x) = x + \Delta\tau \mathbf{C}(f^n(x) - x)\,, \tag{27.14}$$

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 27.2 the method works well for $n \leq 20$. As in (27.12), the multipoint shooting method is the method of preference for determining longer cycles. Consider $x = (x_1, x_2, \ldots, x_n)$ and the $nd$-dimensional map

$$x' = f(x) = (f(x_n), f(x_1), \ldots, f(x_{n-1}))\,.$$

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 $\pm 1$ 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 fundamental 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.

## 27.3 Least action method

(P. Dahlqvist)

The methods of sects. 27.1 and 27.2 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 (27.3), to vary.

**Table 27.3:** 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 $\Lambda_p$, 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 $\overline{001011}$ and $\overline{001101}$ are degenerate due to the time reversal symmetry, but are not related by any discrete spatial symmetry. (Computed by P.E. Rosenqvist.)

| $p$ | $\Lambda_p$ | $T_p$ |
|---|---|---|
| 0 | 9.898979485566 | 4.000000000000 |
| 1 | $-1.177145519638\times10^{1}$ | 4.267949192431 |
| 01 | $-1.240948019921\times10^{2}$ | 8.316529485168 |
| 001 | $-1.240542557041\times10^{3}$ | 12.321746616182 |
| 011 | $1.449545074956\times10^{3}$ | 12.580807741032 |
| 0001 | $-1.229570686196\times10^{4}$ | 16.322276474382 |
| 0011 | $1.445997591902\times10^{4}$ | 16.585242906081 |
| 0111 | $-1.707901900894\times10^{4}$ | 16.849071859224 |
| 00001 | $-1.217338387051\times10^{5}$ | 20.322330025739 |
| 00011 | $1.432820951544\times10^{5}$ | 20.585689671758 |
| 00101 | $1.539257907420\times10^{5}$ | 20.638238386018 |
| 00111 | $-1.704107155425\times10^{5}$ | 20.853571517227 |
| 01011 | $-1.799019479426\times10^{5}$ | 20.897369388186 |
| 01111 | $2.010247347433\times10^{5}$ | 21.116994322373 |
| 000001 | $-1.205062923819\times10^{6}$ | 24.322335435738 |
| 000011 | $1.418521622814\times10^{6}$ | 24.585734788507 |
| 000101 | $1.525597448217\times10^{6}$ | 24.638760250323 |
| 000111 | $-1.688624934257\times10^{6}$ | 24.854025100071 |
| 001011 | $-1.796354939785\times10^{6}$ | 24.902167001066 |
| 001101 | $-1.796354939785\times10^{6}$ | 24.902167001066 |
| 001111 | $2.005733106218\times10^{6}$ | 25.121488488111 |
| 010111 | $2.119615015369\times10^{6}$ | 25.165628236279 |
| 011111 | $-2.366378254801\times10^{6}$ | 25.384945785676 |

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.

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. 12.3 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 $s = (s_1, \ldots, s_n)$, that is find the roots of $\partial_i L(s) = 0$. Expand to first order

$$\partial_i L(s_0 + \delta s) = \partial_i L(s_0) + \sum_j \partial_i \partial_j L(s_0) \delta s_j + \ldots$$

[exercise 27.1]

and use $M_{ij}(s_0) = \partial_i \partial_j L(s_0)$ in the $n$-dimensional Newton-Raphson iteration scheme of sect. 12.2.2

$$s_i \mapsto s_i - \sum_j \left( \frac{1}{M(s)} \right)_{ij} \partial_j L(s) \tag{27.15}$$

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 27.2]
[exercise 12.10]

As an example, the short periods and stabilities of 3-disk cycles computed this way are listed table **??**.

## 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 length.

## Commentary

**Remark 27.1** Piecewise linear maps. The Lozi map (3.20) is linear, and 100,000's of cycles can be easily computed by [2x2] matrix multiplication and inversion.

**Remark 27.2** Relaxation method. The relaxation (or gradient) algorithm is one of the methods for solving extremal problems [13]. The method described above was introduced by Biham and Wenzel [1], 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 [2]. The applicability and reliability of the method is discussed in detail by Grassberger, Kantz and Moening [5], who give examples of the ways in which the method fails: (a) it might reach a limit cycle rather than a equilibrium saddle point (that can be remedied by the complex Biham-Wenzel algorithm [2]) (b) different symbol sequences can converge to the same cycle (i.e., more refined initial conditions might be needed). Furthermore, Hansen (ref. [7] and chapter 4. of ref. [8]) 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. 12.3, 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 27.3** Hybrid Newton-Raphson/relaxation methods. The method discussed in sect. 27.2 was introduced by Schmelcher *et al* [9]. The method was extended to flows by means of the Poincaré surface of section technique in ref. [10]. It is also possible to combine the Newton-Raphson method and (27.13) in the construction of a transformed map [14]. 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 27.4** Relation to the Smale horseshoe symbolic dynamics. For a complete horseshoe Hénon repeller ($a$ sufficiently large), such as the one given in figure 27.2, the signs $\sigma_i \in \{1, -1\}$ are in a 1-to-1 correspondence with the Smale horsheshoe symbolic dynamics $s_i \in \{0, 1\}$:

$$s_i = \begin{cases} 0 & \text{if } \sigma_i = -1\,, \quad x_i < 0 \\ 1 & \text{if } \sigma_i = +1\,, \quad x_i > 0 \end{cases} \quad . \tag{27.16}$$

For arbitrary parameter values with a finite subshift symbolic dynamics or with arbitrarily complicated pruning, the relation of sign sequences $\{\sigma_1, \sigma_2, \cdots, \sigma_n\}$ to the itineraries $\{s_1, s_2, \cdots, s_n\}$ can be much subtler; this is discussed in ref. [5].

**Remark 27.5** Ikeda map. Ikeda map (27.11) was introduced in ref. [12] is a model which exhibits complex dynamics observed in nonlinear optical ring cavities.

**Remark 27.6** 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 surface of 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. [15, 16].

**Remark 27.7** Stability ordering. The parameter $\Delta\tau$ in (27.13) 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 fundamental 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 [11]. 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 stability eigenvalues of the fixed points stabilized in the course of decreasing values for $\Delta\tau$. As explained in sect. 18.5, 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 [13, 14, 12], and shown to yield good results in practical applications.

[section 18.5]

**Remark 27.8** Action extremization method. The action extremization (sect. 27.3) 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 [17, 18]. 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 [20] 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 [19].

A fictitious time gradient flow similar to the one discussed here in sect. 27.1 was introduced by Anegent [21] for twist maps, and used by Gole [22] 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 Tompaidis [23, 24] 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. [25]).

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 [23] do find it useful to minimize instead $S\bar{S}$, analogous to our cost function (27.3).

## Exercises

27.1. **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 **??**. (One such method is given in sect. 27.3.) After that check the accuracy of the computed orbits by iterating them forward with your simulator. What is your error $|f^{T_p}(x) - x|$?

27.2. **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 $R : a = 2.04821419\ldots$ families of cycles are pruned, i.e. some of the minimal length trajectories are blocked by intervening disks.

27.3. **Cycles of the Hénon map.** Apply the method of sect. 27.1 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 (13.1), or as the zero of a truncated topological zeta function (13.21). Do your cycles agree with the cycles listed in table **??**?

## References

[27.1] O. Biham and W. Wenzel, "Characterization of unstable periodic orbits in chaotic attractors and repellers," *Phys. Rev. Lett.* **63**, 819 (1989).

[27.2] O. Biham and W. Wenzel, *Phys. Rev.* **A 42**, 4639 (1990).

[27.3] P. Grassberger and H. Kantz, "Generating partitions for the dissipative Hénon map," *Phys. Lett. A* **113**, 235 (1985).

[27.4] H. Kantz and P. Grassberger, *Physica* **17D**, 75 (1985).

[27.5] P. Grassberger, H. Kantz, and U. Moenig. "On the symbolic dynamics of the Hénon map," *J. Phys.* **A 43**, 5217 (1989).

[27.6] M. Eisele, "Comparison of several generating partitions of the Hnon map," *J. Phys.* **A 32**, 1533 (1999).

[27.7] K.T. Hansen, "Remarks on the symbolic dynamics for the Hénon map," *Phys. Lett.* **A 165**, 100 (1992).

[27.8] D. Sterling and J.D. Meiss, "Computing periodic orbits using the anti-integrable limit," *Phys. Lett. A* **241**, 46 (1998); arXiv:chao-dyn/9802014.

[27.9] P. Schmelcher and F.K. Diakonos, *Phys. Rev. Lett.* **78**, 4733 (1997); *Phys. Rev. E* **57**, 2739 (1998).

[27.10] D. Pingel, P. Schmelcher and F.K. Diakonos, O. Biham, *Phys. Rev. E* **64**, 026214 (2001).

[27.11] F. K. Diakonos, P. Schmelcher, O. Biham, *Phys. Rev. Lett.* **81**, 4349 (1998).

[27.12] K. Ikeda, *Opt. Commun.* **30**, 257 (1979).

[27.13] F. Stummel and K. Hainer, *Praktische Mathematik* (Teubner, Stuttgart 1982).

[27.14] R.L. Davidchack and Y.C. Lai, *Phys. Rev. E* **60**, 6172 (1999).

[27.15] P. Cvitanović and Y. Lan, "Turbulent fields and their recurrences," in N. Antoniou, ed., *Proceed. of 10. Intern. Workshop on Multiparticle Production: Correlations and Fluctuations in QCD* (World Scientific, Singapore 2003). `nlin.CD/0308006`

[27.16] Y. Lan and P. Cvitanović, "Variational method for finding periodic orbits in a general flow," *Phys. Rev.* **E 69** 016217 (2004), `nlin.CD/0308008`

[27.17] M. Sieber and F. Steiner, " Quantum Chaos in the Hyperbola Billiard," *Phys. Lett.* **A 148**, 415 (1990).

[27.18] M. Sieber, *The Hyperbola Billiard: A Model for the Semiclassical Quantization of Chaotic Systems*, Ph.D. thesis (Hamburg 1991); DESY report 91-030.

[27.19] H. T. Kook and J. D. Meiss, "Periodic orbits for reversible symplectic mappings," *Physica D* **35**, 65 (1989).

[27.20] J.N. Mather, "Variational construction of orbits of twist difeomorphisms," J. Amer. Math. Soc. **4** 207 (1991).

[27.21] S. B. Angenent, "The periodic orbits of an area preserving twist-map," *Comm. Math. Phys.* **115**, 353 (1988).

[27.22] C. Golé, "A new proof of the Aubry-Mather's theorem," *Math. Z.* **210**, 441 (1992).

[27.23] 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 (1992).

[27.24] S. Tompaidis, "Numerical Study of Invariant Sets of a Quasi-periodic Perturbation of a Symplectic Map," *Experimental Mathematics* **5**, 211 (1996).

[27.25] R. de la Llave, *Variational methods for quasiperiodic solutions of partial differential equations*, `mp_arc 00-56`.

# Chapter 28

# 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 intuition is much facilitated by approximating circle maps by number-theoretic

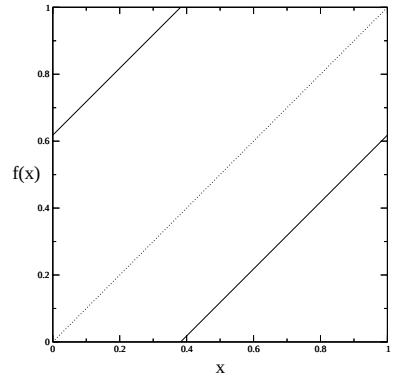



**Figure 28.1:** Unperturbed circle map ($k = 0$ in (28.1)) with golden mean rotation number.

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.

## 28.1 Mode locking

The simplest way of modeling a nonlinearly perturbed rotation on a circle is by 1-dimensional circle maps $x \to x' = f(x)$, restricted to the one dimensional torus, such as the *sine map*

$$x_{n+1} = f(x_n) = x_n + \Omega - \frac{k}{2\pi}\sin(2\pi x_n) \qquad \text{mod } 1 \ . \tag{28.1}$$

$f(x)$ 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 $\Omega$ is the *bare* frequency.

The state space of this map, the unit interval, can be thought of as the elementary cell of the map

$$\hat{x}_{n+1} = \hat{f}(\hat{x}_n) = \hat{x}_n + \Omega - \frac{k}{2\pi}\sin(2\pi\hat{x}_n) \ . \tag{28.2}$$

where $\hat{}$ is used in the same sense as in chapter 24.

The winding number is defined as

$$W(k,\Omega) = \lim_{n\to\infty}(\hat{x}_n - \hat{x}_0)/n. \tag{28.3}$$

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 28.1

$$x_{n+1} = x_n + \Omega \qquad \text{mod } 1 \ , \tag{28.4}$$

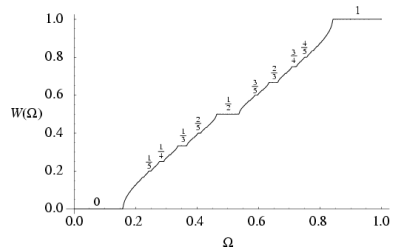



**Figure 28.2:** The critical circle map ($k = 1$ in (28.1)) devil's staircase [3]; the winding number $W$ as function of the parameter $\Omega$.

and the rotation number is given by the parameter $\Omega$.

$$W(k = 0, \Omega) = \Omega \ .$$

For given values of $\Omega$ 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 $Q$

$$\hat{f}^Q(\hat{x}_i) = \hat{x}_i + P, \quad i = 0, 1, 2, \cdots, Q - 1 \ .$$

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 $\Omega$ 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 28.1]

$$\Delta_{P/Q} = Q^{-2\mu_{P/Q}} = \Omega^{right}_{P/Q} - \Omega^{left}_{P/Q} \ . \tag{28.5}$$

where $\Omega^{right}_{P/Q}$ ($\Omega^{left}_{P/Q}$) denote the biggest (smallest) value of $\Omega$ for which $W(k, \Omega) = P/Q$. Parametrizing mode lockings by the exponent $\mu$ rather than the width $\Delta$ will be convenient for description of the distribution of the mode-locking widths, as the exponents $\mu$ 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 $\Omega^{right}_{P/Q}$, $\Omega^{left}_{P/Q}$ endpoints of (28.5)). For the shift map (28.4), the stability intervals are shrunk to points. As $\Omega$ 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 $\Omega$ is a convenient visualization of the mode-locking structure of circle maps. It yields a monotonic "devil's staircase" of figure 28.2 whose self-similar structure we are to unravel. Circle maps with zero slope at the inflection point $x_c$

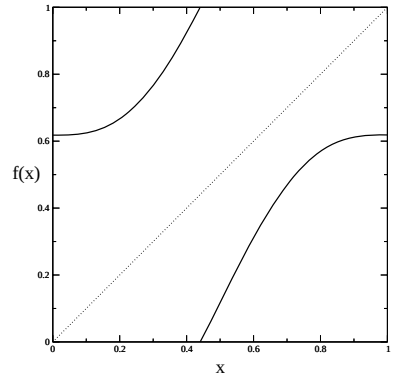



**Figure 28.3:** Critical circle map ($k = 1$ in (28.1)) with golden mean bare rotation number.

(see figure 28.3)

$$f'(x_c) = 0\,, \qquad f''(x_c) = 0$$

($k = 1$, $x_c = 0$ in (28.1)) 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.

### 28.1.1 Hierarchical partitions of the rationals

Intuitively, the longer the cycle, the finer the tuning of the parameter $\Omega$ required to attain it; given finite time and resolution, we expect to be able to resolve cycles up to some maximal length $Q$. This is the physical motivation for partitioning mode lockings into sets of cycle length 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 $0 < P_i \le Q_i \le Q$ 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

$$\Phi(Q) = \sum_{n=1}^{Q} \phi(Q) = \frac{3Q^2}{\pi^2} + O(Q \ln Q). \tag{28.6}$$

Here the Euler function $\phi(Q)$ is the number of integers not exceeding and relatively prime to $Q$. For example, $\phi(1) = 1$, $\phi(2) = 1$, $\phi(3) = 2, \ldots, \phi(12) = 4, \phi(13) = 12, \ldots$

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 $\Delta_{[1]}, \Delta_{[2]}, \cdots, \Delta_{[a_1]}, \cdots$ mode-lockings; their complement are the *covering* intervals $\ell_1, \ell_2, \ldots, \ell_{a_1}, \ldots$ which contain all windings, rational and irrational, whose continued fraction expansion starts with $[a_1, \ldots]$ and is of length at least 2. The second level is obtained by deleting $\Delta_{[1,2]}, \Delta_{[1,3]}, \cdots, \Delta_{[2,2]}, \Delta_{[2,3]}, \cdots, \Delta_{[n,m]}, \cdots$ and so on.

The $n$th level continued fraction partition $\mathcal{S}_n = \{a_1 a_2 \cdots a_n\}$ 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:

$$\frac{P_i}{Q_i} = [a_1, a_2, \cdots, a_n] = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ldots \cfrac{1}{a_n}}}$$

The object of interest, the set of the irrational winding numbers, is in this partitioning labeled by $\mathcal{S}_\infty = \{a_1 a_2 a_3 \cdots\}$, $a_k \in Z^+$, 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

$$T(x) = \begin{array}{ll} \dfrac{1}{x} - \left[\dfrac{1}{x}\right] & x \neq 0 \\ 0\ , & x = 0 \end{array} \tag{28.7}$$

($[\cdots]$ denotes the integer part) acts as a shift on the continued fraction representation of numbers on the unit interval

$$x = [a_1, a_2, a_3, \ldots] \rightarrow T(x) = [a_2, a_3, \ldots]\ . \tag{28.8}$$

into the "mother" interval $\ell_{a_2 a_3 \ldots}$.

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 halfways between any two large stability intervals (such as $1/2$ and $1/3$) in the devil's staircase of figure 28.2 there is the next largest stability interval (such as $2/5$). The winding number of this interval is given by the Farey mediant $(P+P')/(Q+Q')$ 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 length $Q+Q'$. 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.

We define the $n$th *Farey tree level* $T_n$ *as the monotonically increasing sequence of those continued fractions* $[a_1, a_2, \ldots, a_k]$ *whose entries* $a_i \geq 1$, $i = 1, 2, \ldots, k-1$, $a_k \geq 2$, *add up to* $\sum_{i=1}^{k} a_i = n + 2$. For example

$$T_2 = \{[4], [2,2], [1,1,2], [1,3]\} = \left(\frac{1}{4}, \frac{1}{5}, \frac{3}{5}, \frac{3}{4}\right). \tag{28.9}$$

The number of terms in $T_n$ is $2^n$. Each rational in $T_{n-1}$ has two "daughters" in $T_n$, given by

$$\begin{array}{ccc} & [\cdots, a] & \\ [\cdots, a-1, 2] & & [\cdots, a+1] \end{array}$$

Iteration of this rule places all rationals on a binary tree, labeling each by a unique binary label, figure 28.4.

The smallest and the largest denominator in $T_n$ are respectively given by

$$[n-2] = \frac{1}{n-2}, \qquad [1, 1, \ldots, 1, 2] = \frac{F_{n+1}}{F_{n+2}} \propto \rho^n , \tag{28.10}$$

where the Fibonacci numbers $F_n$ are defined by $F_{n+1} = F_n + F_{n-1}$; $F_0 = 0$, $F_1 = 1$, and $\rho$ is the golden mean ratio

$$\rho = \frac{1+\sqrt{5}}{2} = 1.61803\ldots \tag{28.11}$$

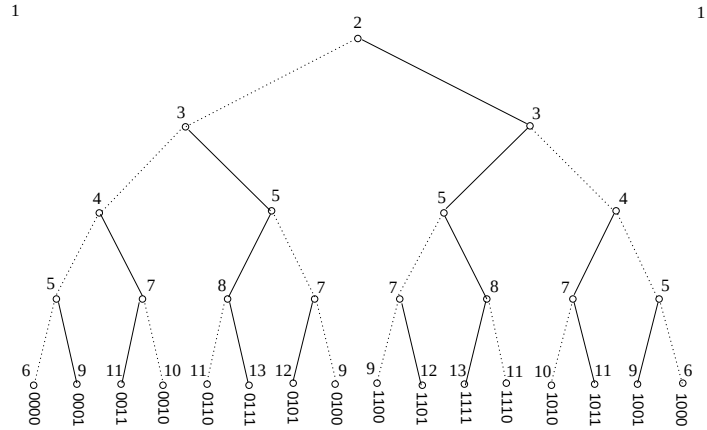



**Figure 28.4:** Farey tree: alternating binary ordered labeling of all Farey denominators on the $n$th Farey tree level.

Note the enormous spread in the cycle lengths on the same level of the Farey tree: $n \leq Q \leq \rho^n$. The cycles whose length grows only as a power of the Farey tree level will cause strong non-hyperbolic 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.

## 28.2 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 $t = n_1, n_2, n_3, \cdots$ 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 $\alpha$), in the parameter space (by shifting the parameter origin onto the border and rescaling by a factor $\delta$), and by replacing the initial map $f$ by the $n$th iterate $f^n$ restricted to the magnified neighborhood

$$f_p(x) \to R f_p(x) = \alpha f^n_{p/\delta}(x/\alpha)$$

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 fixpoint 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 (28.10). 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 (28.38); it maps a circle map (represented as a pair of

functions $(g, f)$, of winding number $[a, b, c, \ldots]$ into a rescaled map of winding number $[b, c, \ldots]$:

$$R_a \begin{pmatrix} g \\ f \end{pmatrix} = \begin{pmatrix} \alpha g^{a-1} \circ f \circ \alpha^{-1} \\ \alpha g^{a-1} \circ f \circ g \circ \alpha^{-1} \end{pmatrix}, \tag{28.12}$$

Acting on a map with winding number $[a, a, a, \ldots]$, $R_a$ returns a map with the same winding number $[a, a, \ldots]$, so the fixed point of $R_a$ has a quadratic irrational winding number $W = [a, a, a, \ldots]$. This fixed point has a single expanding eigenvalue $\delta_a$. Similarly, the renormalization transformation $R_{a_p} \ldots R_{a_2} R_{a_1} \equiv R_{a_1 a_2 \ldots a_p}$ has a fixed point of winding number $W_p = [a_1, a_2, \ldots, a_{n_p}, a_1, a_2, \ldots]$, with a single expanding eigenvalue $\delta_p$.

For short repeating blocks, $\delta$ 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 $\Omega_r$ be the parameter for which the map (28.1) has a superstable cycle of rotation number $P_r/Q_r = [p, p, \ldots, p]$. The $\delta_p$ can then be estimated by extrapolating from

$$\Omega_r - \Omega_{r+1} \propto \delta_p^{-r}. \tag{28.13}$$

What this means is that the "devil's staircase" of figure 28.2 is self-similar under magnification by factor $\delta_p$ 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 (28.13) is the sequence of Fibonacci number continued fraction approximants to the golden mean winding number $W = [1, 1, 1, ...] = (\sqrt{5} - 1)/2$.

When global problems are considered, it is useful to have at least and idea on extemal scaling laws for mode–lockings. This is achieved, in a first analysis, by fixing the cycle length $Q$ and describing the range of possible asymptotics.

For a given cycle length $Q$, it is found that the *narrowest* interval shrinks with a power law

$$\Delta_{1/Q} \propto Q^{-3} \tag{28.14}$$

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

$$\Delta_{P/Q} \propto Q^{-2\mu_1} \tag{28.15}$$

where $P = F_{n-1}$, $Q = F_n$ and $\mu_1$ is related to the universal Shenker number $\delta_1$ (28.13) and the golden mean (28.11) by

$$\mu_1 = \frac{\ln|\delta_1|}{2\ln\rho} = 1.08218\ldots \tag{28.16}$$

The closeness of $\mu_1$ 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:

$$3/2 > \mu_{m/n} \geq 1.08218\cdots. \tag{28.17}$$

## 28.3 Global theory: Thermodynamic averaging

Consider the following average over mode-locking intervals (28.5):

$$\Omega(\tau) = \sum_{Q=1}^{\infty} \sum_{(P|Q)=1} \Delta_{P/Q}^{-\tau}. \tag{28.18}$$

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 length $Q$, with winding number $P/Q$.

The qualitative behavior of (28.18) is easy to pin down. For sufficiently negative $\tau$, the sum is convergent; in particular, for $\tau = -1$, $\Omega(-1) = 1$, as for the critical circle maps the mode-lockings fill the entire $\Omega$ range [11]. However, as $\tau$ 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 $\tau$ the sum diverges. This occurs for $\tau < 0$, as $\Omega(0)$ equals the number of all rationals and is clearly divergent.

The sum (28.18) 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:

$$\bar{Z}_n(\tau) = \sum_{i\in\mathcal{S}_n} \Delta_i^{-\tau}. \tag{28.19}$$

The original sum (28.18) 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 (28.18):

$$\Omega(q,\tau) = \sum_{Q=1}^{\infty} \sum_{(P|Q)=1} e^{-q\nu_{P/Q}} Q^{2\tau\mu_{P/Q}} \ . \tag{28.20}$$

The sum (28.18) 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 of the $\nu_{P/Q}$ hierarchy that we consider here correspond respectively to the Farey series partitioning

$$\Omega(q,\tau) = \sum_{Q=1}^{\infty} \Phi(Q)^{-q} \sum_{(P|Q)=1} Q^{2\tau\mu_{P/Q}} \ , \tag{28.21}$$

the continued fraction partitioning

$$\Omega(q,\tau) = \sum_{n=1}^{\infty} e^{-qn} \sum_{[a_1,\ldots,a_n]} Q^{2\tau\mu_{[a_1,\ldots,a_n]}} \ , \tag{28.22}$$

and the Farey tree partitioning

$$\Omega(q,\tau) = \sum_{k=n}^{\infty} 2^{-qn} \sum_{i=1}^{2^n} Q_i^{2\tau\mu_i} \ , \quad Q_i/P_i \in T_n \ . \tag{28.23}$$

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.

## 28.4 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

$$\ell_i = \Omega^{min}_{P_r/Q_r} - \Omega^{max}_{P_l/Q_l} \ . \tag{28.24}$$

Here $\Omega^{min}_{P_r/Q_r}$ ($\Omega^{max}_{P_l/Q_l}$) are respectively the lower (upper) edges of the mode-locking intervals $\Delta_{P_r/Q_r}$ ($\Delta_{P_l/Q_l}$) bounding $\ell_i$ and $i$ is a symbolic dynamics label, for example the entries of the continued fraction representation $P/Q = [a_1, a_2, \ldots, a_n]$ of

one of the boundary mode-lockings, $i = a_1 a_2 \cdots a_n$. $\ell_i$ provide a finite cover for the irrational winding set, so one may consider the sum

$$Z_n(\tau) = \sum_{i \in \mathcal{S}_n} \ell_i^{-\tau} \tag{28.25}$$

The value of $-\tau$ for which the $n \to \infty$ limit of the sum (28.25) 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 $\ell_i$ is optimal; otherwise it provides an upper bound to $D_H$. As by construction the $\ell_i$ 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 $D_H = 0.870\ldots$ is a (global) universal number.

[exercise 28.2]

### 28.4.1 The Hausdorff dimension in terms of cycles

Estimating the $n \to \infty$ limit of (28.25) from finite numbers of covering intervals $\ell_i$ 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 (28.4), the covering intervals are generated by iterations of the Farey map (28.37) or the Gauss shift (28.38). The $n$th level sum (28.25) 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 $\ell_i$ by $|df^n/dx|$ evaluated on the $i$th interval. We are thus led to the following determinant

$$\begin{aligned} \det(1 - z\mathcal{L}_\tau) &= \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{z^{rn_p}}{r} \frac{|\Lambda_p^r|^\tau}{1 - 1/\Lambda_p^r}\right) \\ &= \prod_p \prod_{k=0}^{\infty} \left(1 - z^{n_p}|\Lambda_p|^\tau/\Lambda_p^k\right) . \end{aligned} \tag{28.26}$$

The sum (28.25) 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 (28.26); putting all these pieces together, we obtain a pretty formula relating the Hausdorff dimension to the prime cycles of the map $f(x)$:

$$0 = \prod_p \left(1 - 1/|\Lambda_p|^{D_H}\right) . \tag{28.27}$$

**Table 28.1:** Shenker's $\delta_p$ for a few periodic continued fractions, from ref. [1].

| $p$ | $\delta_p$ |
|---|---|
| [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 |

For the Gauss shift (28.38) the stabilities of periodic cycles are available analytically, as roots of quadratic equations: For example, the $x_a$ fixed points (quadratic irrationals with $x_a = [a, a, a \ldots]$ infinitely repeating continued fraction expansion) are given by

$$x_a = \frac{-a + \sqrt{a^2+4}}{2}, \qquad \Lambda_a = -\left(\frac{a + \sqrt{a^2+4}}{2}\right)^2 \tag{28.28}$$

and the $x_{ab} = [a, b, a, b, a, b, \ldots]$ 2–cycles are given by

$$\begin{aligned} x_{ab} &= \frac{-ab + \sqrt{(ab)^2 + 4ab}}{2b} \\ \Lambda_{ab} &= (x_{ab}x_{ba})^{-2} = \left(\frac{ab + 2 + \sqrt{ab(ab+4)}}{2}\right)^2 \end{aligned} \tag{28.29}$$

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 (28.27) merely a very convoluted way to compute the number 1? Possibly so, but once the meaning of (28.27) has been grasped, the corresponding formula for the *critical* circle maps follows immediately:

$$0 = \prod_p \left(1 - 1/|\delta_p|^{D_H}\right) \ . \tag{28.30}$$

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 (28.30) are rather involved, as they require a heavy computational effort to extract Shenker's scaling $\delta_p$ 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 **??** we give a small selection of computed Shenker's scalings.

## 28.5 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 chapter 22, by looking at the free energy.

Consider the Farey tree partition sum (28.23): the narrowest mode-locked interval (28.15) at the $n$th level of the Farey tree partition sum (28.23) is the golden mean interval

$$\Delta_{F_{n-1}/F_n} \propto |\delta_1|^{-n}. \tag{28.31}$$

It shrinks exponentially, and for $\tau$ positive and large it dominates $q(\tau)$ and bounds $dq(\tau)/d\tau$:

$$q'_{max} = \frac{\ln|\delta_1|}{\ln 2} = 1.502642\ldots \tag{28.32}$$

However, for $\tau$ large and negative, $q(\tau)$ is dominated by the interval (28.14) which shrinks only harmonically, and $q(\tau)$ approaches 0 as

$$\frac{q(\tau)}{\tau} = \frac{3\ln n}{n\ln 2} \to 0. \tag{28.33}$$

So for finite n, $q_n(\tau)$ crosses the $\tau$ axis at $-\tau = D_n$, but in the $n \to \infty$ limit, the $q(\tau)$ function exhibits a phase transition; $q(\tau) = 0$ for $\tau < -D_H$, but is a non-trivial function of $\tau$ 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 $Q^{-2}$:

$$Z_n(\tau) = \sum_{i=1}^{2^n} Q_i^{2\tau}. \tag{28.34}$$

Here $Q_i$ is the denominator of the $i$th Farey rational $P_i/Q_i$. For example (see figure 28.4),

$$Z_2(1/2) = 4 + 5 + 5 + 4.$$

By the annihilation property (28.38) of the Gauss shift on rationals, the $n$th Farey level sum $Z_n(-1)$ can be written as the integral

$$Z_n(-1) = \int dx\, \delta(f^n(x)) = \sum 1/|f'_{a_1\ldots a_k}(0)|\,,$$

| $\tau/2$ | $Z_n(\tau/2)/Z_{n-1}(\tau/2)$ |
|---|---|
| 0 | 2 |
| 1 | 3 |
| 2 | $(5+\sqrt{17})/2$ |
| 3 | 7 |
| 4 | $(5+\sqrt{17})/2$ |
| 5 | $7+4\sqrt{6}$ |
| 6 | $26.20249\ldots$ |

**Table 28.2:** 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 $a_1 + \ldots + a_k = n+2$. It is easily checked that $f'_{a_1\ldots a_k}(0) = (-1)^k Q^2_{[a_1,\ldots,a_k]}$, 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 (28.26). 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 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, $Z_n(3/2) = \sum_i Q^3 = 54\cdot 7^{n-1}$. A few of these "sum rules" are listed in the table 28.2, they are consequence of the fact that the denominators on a given level are Farey sums of denominators on preceding levels.

[exercise 28.3]

A bound on $D_H$ can be obtained by approximating (28.34) by

$$Z_n(\tau) \;=\; n^{2\tau} \;+\; 2^n\rho^{2n\tau}. \tag{28.35}$$

In this approximation we have replaced all $\ell_{P/Q}$, except the widest interval $\ell_{1/n}$, by the narrowest interval $\ell_{F_{n-1}/F_n}$ (see (28.15)). The crossover from the harmonic dominated to the golden mean dominated behavior occurs at the $\tau$ value for which the two terms in (28.35) contribute equally:

$$D_n \;=\; \hat{D} \;+\; O\left(\frac{\ln n}{n}\right), \quad \hat{D} \;=\; \frac{\ln 2}{2\ln\rho} = .72\ldots \tag{28.36}$$

For negative $\tau$ the sum (28.35) is the lower bound on the sum (28.25) , 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 webs. Thus the considerations of chapter 23 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 28.1** The physics of circle maps. Mode–locking phenomenology is reviewed in ref. [5], a more theoretically oriented discussion is contained in ref. [3]. While representative of dissipative systems we may also consider circle mapsas a crude approximation to Hamiltonian local dynamics: a typical island of stability in a Hamiltonian 2-$d$ 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 $\Omega$, and the angular motion can be modelled by a map periodic in the angular variable [8, 9]. 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 [5, 8] and non-hyperbolicity 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 webs, development of good computational techniques is here of utmost practical importance.

**Remark 28.2** Critical mode–locking set The fact that mode-lockings completely fill the unit interval at the critical point has been proposed in refs. [**?**, 10]. The proof that the set of irrational windings is of zero Lebesgue measure in given in ref. [11].

**Remark 28.3** 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 $Q \to \infty$ 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 (28.25) of any methods used so far; for example the computation in ref. [12] for critical sine map (28.1), 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 28.4** Farey tree presentation function. The Farey tree rationals can be generated by backward iterates of 1/2 by the Farey presentation function [ 13]:

$$\begin{array}{rcll} f_0(x) & = & x/(1-x) & 0 \le x < 1/2 \\ f_1(x) & = & (1-x)/x & 1/2 < x \le 1 \ . \end{array} \tag{28.37}$$

The Gauss shift (28.7) corresponds to replacing the binary Farey presentation function branch $f_0$ in (28.37) by an infinity of branches

$$\begin{array}{rcll} f_a(x) & = & f_1 \circ f_0^{(a-1)}(x) = \dfrac{1}{x} - a, & \dfrac{1}{a-1} < x \le \dfrac{1}{a}, \\ f_{ab\cdots c}(x) & = & f_c \circ \cdot \circ f_b \circ f_a(x) \ . & \end{array} \tag{28.38}$$

A rational $x = [a_1, a_2, \ldots, a_k]$ is annihilated by the $k$th iterate of the Gauss shift, $f_{a_1 a_2 \cdots a_k}(x) = 0$. 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 ( 28.37) 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. [4].

An alternative labeling of Farey denominators has been introduced by Knauf [ 6] 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 28.5** Circle map renormalization The idea underlying golden mean renormalization goes back to Shenker [9]. A renormalization group procedure was formulated in refs. [7, 14], where moreover the uniqueness of the relevant eigenvalue is claimed. This statement has been confirmed by a computer–assisted proof [ 15], and in the following we will always assume it. There are a number of experimental evidences for local universality, see refs. [16, 17].

On the other side of the scaling tale, the power law scaling for harmonic fractions (discussed in refs. [2, **?**, 4]) is derived by methods akin to those used in describing intermittency [21]: $1/Q$ 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 28.6** 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 [ 22, 23] . 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 [24]. 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:

$$\delta_i = \frac{i}{\Phi(Q)} - \frac{P_i}{Q_i}, \quad i = 1, 2, \cdots, \Phi(Q)$$

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 [25, 22, 23], the Riemann hypothesis is equivalent to the statement that

$$\sum_{Q_i \le Q} |\delta_i| = o(Q^{\frac{1}{2}+\epsilon})$$

for all $\epsilon$ as $Q \to \infty$. The mode-lockings $\Delta_{P/Q}$ contain the necessary information for constructing the partition of the unit interval into the $\ell_i$ covers, and therefore implicitly contain the $\delta_i$ 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 28.7** Farey tree partitioning. The Farey tree partitioning was introduced in refs. [26, 27, 4] and its thermodynamics is discussed in detail in refs. [12, 13]. 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 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 28.8** Local and global universality. Numerical evidences for global universal behavior have been presented in ref. [3]. The question was reexamined in ref. [12], 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 (28.30), linking local to global behavior, was proposed in ref. [1].

The derivation of (28.30) relies only on the following aspects of the "hyperbolicity conjecture" of refs. [4, 18, 19, 20]:

1. *limits* for Shenker $\delta$'s *exist* and are universal. This should follow from the renormalization theory developed in refs. [7, 14, 15], though a general proof is still lacking.
2. $\delta_p$ grow *exponentially* with $n_p$, the length of the continued fraction block $p$.
3. $\delta_p$ for $p = a_1 a_2 \ldots n$ with a large continued fraction entry $n$ grows as a *power* of $n$. According to (28.14), $\lim_{n\to\infty} \delta_p \propto n^3$. In the calculation of ref. [1] the explicit values of the asymptotic exponents and prefactors were not used, only the assumption that the growth of $\delta_p$ with $n$ is not slower than a power of $n$.

**Remark 28.9** Farey model. The Farey model (28.33) has been proposed in ref. [12]; 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 (28.26).

**Remark 28.10** Symbolic dynamics for Hamiltonian rotational orbits. The rotational codes of ref. [6] are closely related to those for maps with a natural angle variable, for example for circle maps [34, 36] and cat maps [37]. Ref. [6] also offers a systematic rule for obtaining the symbolic codes of "islands around islands" rotational orbits [39]. 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. [42, 40]; 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 [39, 41].

## Exercises

28.1. **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 $k$.

28.2. **Bounds on Hausdorff dimension.** By making use of the bounds (28.17) show that the Hausdorff dimension for critical mode lockings may be bounded by

$$2/3 \leq D_H \leq .9240\ldots$$

28.3. **Farey model sum rules.** Verify the sum rules reported in table 28.2. 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, see ref. [28].

28.4. **Metric entropy of the Gauss shift.** Check that the Lyapunov exponent of the Gauss map (28.7) is given by $\pi^2/6\ln 2$. This result has been claimed to be relevant in the discussion of "mixmaster" cosmologies, see ref. [30].

28.5. **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 [31].

28.6. **Hitting condition.** Prove (S.39). Hint: together with the real trajectory consider the line passing through the starting point, with polar angle $\theta_{m,n}$: then draw the perpendiculars to the actual trajectory, passing through the center of the $(0,0)$ and $(m,n)$ disks.

28.7. $j_n$ **and** $\alpha_{cr}$**.** Look at the integration region and how it scales by plotting it for increasing values of $n$.

28.8. **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, 33].

28.9. **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 $g(x) = 1/x - [1/x]$.

28.10. **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 righmost, 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. [17].

## References

[28.1] P. Cvitanović, G.H. Gunaratne and M. Vinson, *Nonlinearity* **3** (1990)

[28.2] K. Kaneko, *Prog. Theor. Phys.* **68**, 669 (1982); **69**, 403 (1983); **69**, 1427 (1983)

[28.3] M.H. Jensen, P. Bak, T. Bohr, *Phys. Rev. Lett.* **50**, 1637 (1983); *Phys. Rev.* **A 30**, 1960 (1984); P. Bak, T. Bohr and M.H. Jensen, *Physica Scripta* **T9**, 50 (1985)

[28.4] P. Cvitanović, B. Shraiman and B. Söderberg, *Physica Scripta* **32**, 263 (1985).

[28.5] J.A. Glazier and A. Libchaber, *IEEE Trans. Circ. Syst.*, **35**, 790 (1988)

[28.6] A. Knauf, "On a ferromagnetic spin chain," *Commun. Math. Phys.* **153**, 77 (1993).

[28.7] M.J. Feigenbaum, L.P. Kadanoff, S.J. Shenker, *Physica* **5D**, 370 (1982)

[28.8] S.J. Shenker and L.P. Kadanoff, *J. Stat. Phys.* **27**, 631 (1982)

[28.9] S.J. Shenker, *Physica* **5D**, 405 (1982)

[28.10] O.E. Lanford, *Physica* **14D**, 403 (1985)

[28.11] G. Swiatek, *Commun. Math. Phys.* **119**, 109 (1988)

[28.12] R. Artuso, P. Cvitanović and B.G. Kenny, *Phys. Rev.* **A39**, 268 (1989); P. Cvitanović, in R. Gilmore (ed), *Proceedings of the XV International Colloquium on Group Theoretical Methods in Physics*, (World Scientific, Singapore, 1987)

[28.13] M.J. Feigenbaum, *J.Stat.Phys.* **52**, 527 (1988)

[28.14] S. Ostlund, D.A. Rand, J. Sethna and E. Siggia, *Physica* **D 8**, 303 (1983).

[28.15] B.D. Mestel, Ph.D. Thesis (U. of Warwick 1985).

[28.16] J. Stavans, F. Heslot and A. Libchaber, *Phys. Rev. Lett.* **55**, 569 (1985)

[28.17] E.G. Gwinn and R.M. Westervelt, *Phys. Rev. Lett.* **59**, 157 (1987)

[28.18] O.E. Lanford, in M. Mebkhout and R. Sénéor, eds., *Proc. 1986 IAMP Conference in Mathematical Physics* (World Scientific, Singapore 1987); D.A. Rand, *Proc. R. Soc. London* **A 413**, 45 (1987); *Nonlinearity* **1**, 78 (1988)

[28.19] S.-H. Kim and S. Ostlund, *Physica* **D 39**, 365, (1989)

[28.20] M.J. Feigenbaum, *Nonlinearity* **1**, 577 (1988)

[28.21] Y. Pomeau and P. Manneville, *Commun. Math. Phys.* **74**, 189 (1980); P. Manneville, *J. Phys.* (Paris) **41**, 1235 (1980)

[28.22] H.M. Edwards, *Riemann's Zeta Function* (Academic, New York 1974)

[28.23] E.C. Titchmarsh, *The Theory of Riemann Zeta Function* (Oxford Univ. Press, Oxford 1951); chapter XIV.

[28.24] E.H. Neville, *Roy. Soc. Mathematical Tables* (Cambridge U. Press, Cambridge 1950)

[28.25] J. Franel and E. Landau, *Göttinger Nachr. 198* (1924)

[28.26] G. T. Williams and D. H. Browne, *Amer. Math. Monthly* **54**, 534 (1947)

[28.27] P. Cvitanović and J. Myrheim, *Phys. Lett.* **A94**, 329 (1983); *Commun. Math. Phys.* **121**, 225 (1989)

[28.28] P. Contucci and A. Knauf, *Forum Math.* **9**, 547 (1997)

[28.29] G.H. Hardy and E.M. Wright, *Theory of Numbers* (Oxford Univ. Press, Oxford 1938)

[28.30] A. Csordás and P. Szépfalusy, *Phys. Rev. A* **40**, 2221 (1989) and references therein.

[28.31] P. Dahlqvist, unpublished notes.

[28.32] D. Levin, *Inter. J. Computer Math.* **B3**, 371 (1973).

[28.33] N. Osada, *SIAM J. Numer. Anal.* **27**, 178 (1990).

[28.34] P. Veerman, "Symbolic dynamics and rotation numbers," *Phys. A* **134**, 543 (1986).

[28.35] J.J.P. Veerman and F.M. Tangerman. "Intersection properties of invariant manifolds in certain twist maps," *Comm. Math. Phys.* **139**, 245 (1991).

[28.36] W.-M. Zheng, "Symbolic dynamics for the circle map," *Int. J. Mod. Phys. B* **5**, 481 (1991).

[28.37] I.C. Percival and F. Vivaldi. "A linear code for the sawtooth and cat maps," *Physica D* **27**, 373 (1987).

[28.38] I.C. Percival and F. Vivaldi. "Arithmetical properties of strongly chaotic motion," *Physica D* **25**, 105 (1987).

[28.39] J.D. Meiss, "Class renormalization: Islands around islands," *Phys. Rev. A* **34**, 2375 (1986).

[28.40] V. Afraimovich, A. Maass, and J. Uras. "Symbolic dynamics for sticky sets in Hamiltonian systems," *Nonlinearity* **13**, 617 (2000).

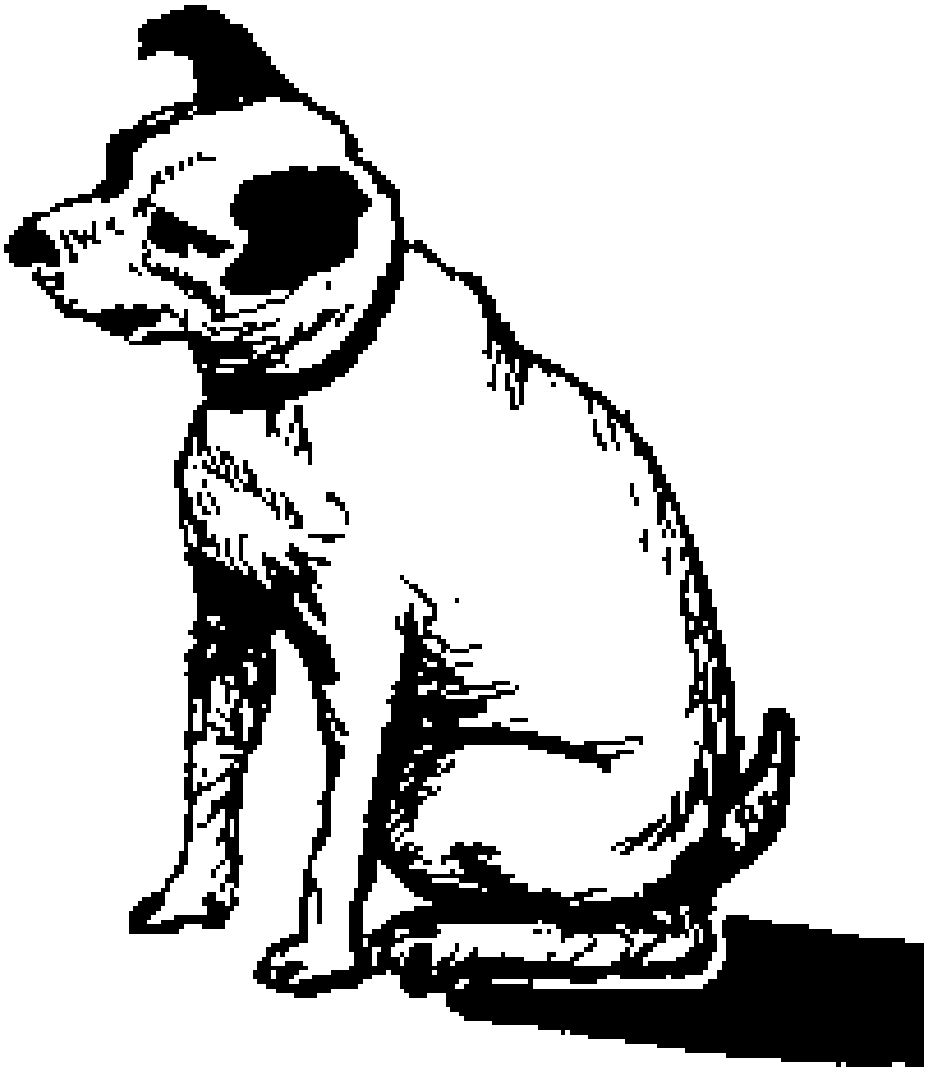
[28.41] J.D. Meiss and E. Ott. "Markov tree model of transport in area preserving maps," *Physica D* **20**, 387 (1986).

[28.42] Y. Aizawa. "Symbolic dynamics approach to the two-D chaos in area-preserving maps," *Prog. Theor. Phys.* **71**, 1419 (1984).

[28.43] M. Yampolsky, "On the eigenvalues of a renormalization operator," *Nonlinearity* **16**, 1565 (2003).

# Chaos: Classical and Quantum

# Volume II: Semiclassical Chaos



**Predrag Cvitanović – Roberto Artuso – Per Dahlqvist – Ronnie Mainieri – Gregor Tanner – Gábor Vattay – Niall Whelan – Andreas Wirzba**

**ChaosBook.org version12, Jun 22 2008** printed June 26, 2008
ChaosBook.org comments to: predrag [snail] nbi.dk

# Chapter 29

# 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)

You have read the first volume of this book. So far, so good – anyone can play a game of classical pinball, and a skilled neuroscientist can poke rat brains. We learned that information about chaotic dynamics can be obtained by calculating spectra of linear operators such as the evolution operator of sect. 15.2 or the associated partial differential equations such as the Liouville equation (14.37). The spectra of these operators can be expressed in terms of periodic orbits of the deterministic dynamics by means of trace formulas and cycle 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 classical flow? The answer is yes, but ... things are becoming somewhat more complicated when studying 2nd or higher order linear PDE's. We can find classical 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 classical dynamics alone, referred to here as the *semiclassical theory*, will not be exact, in contrast to the classical 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 37]

## 29.1 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 has been fully uncovered by Schrödinger in the mid 1920's. Indeed, were physicists of the period as familiar with classical 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 idiosyncracy 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 (1.9) derived in the context of classical chaotic dynamics,

[chapter 17]

$$1/\zeta(z) = \prod_p (1 - t_p) ,$$

also yield excellent estimates of *quantum* resonances, with the quantum amplitude associated with a given cycle approximated semiclassically by the weight

$$t_p = \frac{1}{|\Lambda_p|^{\frac{1}{2}}} e^{\frac{i}{\hbar} S_p - i\pi m_p/2} , \tag{29.1}$$

whose magnitude is the square root of the classical weight (17.10)

$$t_p = \frac{1}{|\Lambda_p|} e^{\beta \cdot A_p - sT_p} ,$$

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 32]

In this approach, the quantal spectra of classically chaotic dynamical systems are determined from the zeros of dynamical zeta functions, defined by cycle expansions of infinite products of form

$$1/\zeta = \prod_p (1 - t_p) = 1 - \sum_f t_f - \sum_k c_k \tag{29.2}$$

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 18 it was shown that for this reason the cycle expansion (29.2) 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 lengths and stabilities of a finite number of shortest periodic orbits, substitutes them into (29.2), 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 classical 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 classical 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 33]

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 18 and 23. 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.

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 classical settings, are also a powerful tool for evaluation of *quantum* resonances of classically 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 34]

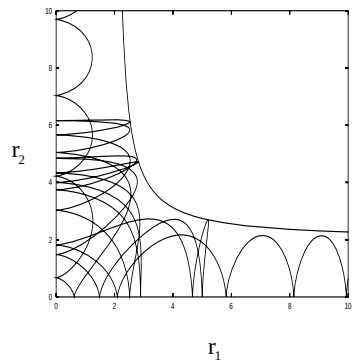
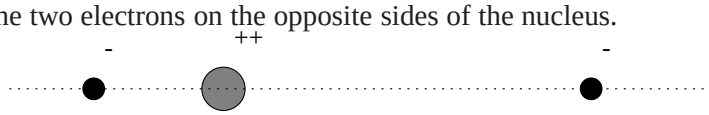



**Figure 29.1:** A typical collinear helium trajectory in the $r_1 - r_2$ plane; the trajectory enters along the $r_1$ axis and escapes to infinity along the $r_2$ axis.

## 29.2 Quantization of helium

Once we have derived the semiclassical weight associated with the periodic orbit $p$ (29.1), 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 classical 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 $\infty$, the helium nucleus charge to 2, the electron charges to -1. The Hamiltonian is [chapter 36]

$$H = \frac{1}{2}p_1^2 + \frac{1}{2}p_2^2 - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_1 + r_2} \,. \tag{29.3}$$

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 29.1, 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 classical collinear helium is also a repeller; almost all of the classical 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 (1.10). 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* (29.1) which is expected to be good only in the classical large energy limit, the eigenenergies are good to 1% all the way down to the ground state. [chapter 36]

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 classical 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 classical 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 33. 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 34, the spectrum of helium in chapter 36, and the incorporation of diffraction effects in chapter 37.

## Commentary

**Remark 29.1** Guide to literature. A key prerequisite to developing any theory of "quantum chaos" is solid understanding of Hamiltonian mechanics. For that, Arnol'd monograph [36] is the essential reference. Ozorio de Almeida's monograph [11] 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, catastrophy theory and torus quantization. The book by Brack and Bhaduri [1] is an excellent introduction to the semiclassical methods. Gutzwiller's monograph [2] is an advanced introduction focusing on chaotic dynamics both in classical 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 text is Reichl's exposition [3].

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 29.2** The dates. Schrödinger's first wave mechanics paper [3] (hydrogen spectrum) was submitted 27 January 1926. Submission date for Madelung's 'quantum theory in hydrodynamical form' paper [2] was 25 October 1926.

## References

[29.1] M. Brack and R.K. Bhaduri, *Semiclassical Physics* (Addison-Wesley, New York 1997).

[29.2] M.C. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer, New York 1990).

[29.3] L.E. Reichl, *The Transition to Chaos in Conservative Classical Systems: Quantum Manifestations* (Springer-Verlag, New York 1992).

[29.4] F. Haake, *Quantum Signatures of Chaos*, 2. edition (Springer-Verlag, New York 2001).

# Chapter 30

# Quantum mechanics, briefly

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

$$i\hbar\frac{\partial}{\partial t}\psi(q,t) = \hat{H}(q, \frac{\hbar}{i}\frac{\partial}{\partial q})\psi(q,t), \tag{30.1}$$

where the Hamilton operator $\hat{H}(q, -i\hbar\partial_q)$ is obtained from the classical Hamiltonian by substitution $p \to -i\hbar\partial_q$. Most of the Hamiltonians we shall consider here are of form

$$H(q,p) = T(p) + V(q)\,, \qquad T(p) = p^2/2m\,, \tag{30.2}$$

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

$$\psi_n(q,t) = e^{-iE_n t/\hbar}\phi_n(q), \tag{30.3}$$

where $E_n$ are the eigenenergies of the time-independent Schrödinger equation

$$\hat{H}\phi(q) = E\phi(q)\,. \tag{30.4}$$

If the kinetic term can be separated out as in (30.2), the time-independent Schrödinger equation

$$-\frac{\hbar^2}{2m}\partial^2\phi(q) + V(q)\phi(q) = E\phi(q) \tag{30.5}$$

can be rewritten in terms of a local wavenumber

$$(\partial^2 + k^2(q))\phi = 0\,, \qquad \hbar^2 k(q) = \sqrt{2m(E - V(q))}\,. \tag{30.6}$$

For bound systems the spectrum is discrete and the eigenfunctions form an orthonormal,

$$\int dq\, \phi_n(q)\phi_m^*(q) = \delta_{nm}\,, \tag{30.7}$$

and complete,

$$\sum_n \phi_n(q)\phi_n^*(q') = \delta(q - q')\,, \tag{30.8}$$

set of functions in a Hilbert space. Here and throughout the text,

$$\int dq = \int dq_1 dq_2 ... dq_D. \tag{30.9}$$

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 34]

A given wave function can be expanded in the energy eigenbasis

$$\psi(q,t) = \sum_n c_n e^{-iE_n t/\hbar}\phi_n(q)\,, \tag{30.10}$$

where the expansion coefficient $c_n$ is given by the projection of the initial wave function $\psi(q,0)$ onto the $n$th eigenstate

$$c_n = \int dq\, \phi_n^*(q)\psi(q,0). \tag{30.11}$$

By substituting (30.11) into (30.10), 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

$$K(q,q',t) = \sum_n \phi_n(q)\, e^{-iE_n t/\hbar}\phi_n^*(q') \tag{30.12}$$

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 $q$

$$K(q, q', t_1 + t_2) = \int dq''\, K(q, q'', t_2) K(q'', q', t_1) \tag{30.13}$$

forward in time, hence the name "propagator." In non-relativistic quantum mechanics the range of $q''$ is infinite, meaning that the wave can propagate at any speed; in relativistic quantum mechanics this is rectified by restricting the propagation to the forward light cone.

Since the propagator is a linear combination of the eigenfunctions of the Schrödinger equation, it also satisfies the Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} K(q, q', t) = \hat{H}(q, \frac{i}{\hbar}\frac{\partial}{\partial q}) K(q, q', t)\,, \tag{30.14}$$

and is thus a wave function defined for $t \geq 0$; from the completeness relation (30.8) we obtain the boundary condition at $t = 0$:

$$\lim_{t \to 0_+} K(q, q', t) = \delta(q - q')\,. \tag{30.15}$$

The propagator thus represents the time evolution of a wave packet which starts out as a configuration space delta-function localized in the point $q'$ at the 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 $\phi_n$ have been determined. With time dependence rendered "trivial," it makes sense to focus on the *Green's function*, the Laplace transformation of the propagator

$$G(q, q', E + i\epsilon) = \frac{1}{i\hbar} \int_0^\infty dt\, e^{\frac{i}{\hbar}Et - \frac{\epsilon}{\hbar}t} K(q, q', t) = \sum_n \frac{\phi_n(q)\phi_n^*(q')}{E - E_n + i\epsilon}\,. \tag{30.16}$$

Here $\epsilon$ 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 the spectrum, one may restrict the considerations to the (formal) trace of the Green's function,

$$\mathrm{tr}\, G(q, q', E) = \int dq\, G(q, q, E) = \sum_n \frac{1}{E - E_n}\,, \tag{30.17}$$

where $E$ is complex, with a positive imaginary part, and we have used the eigenfunction orthonormality (30.7). This trace is formal, since as it stands, the sum

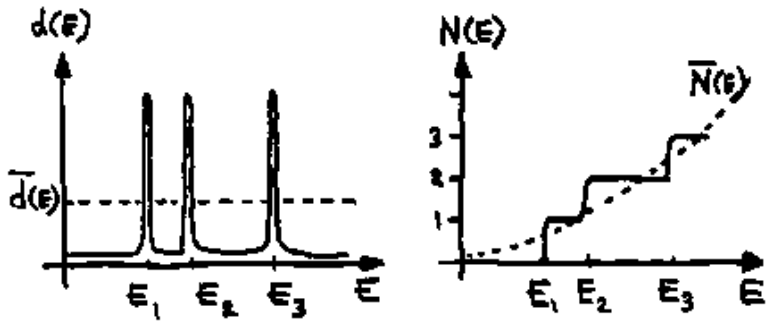



**Figure 30.1:** Schematic picture of **a)** the density of states $d(E)$, and **b)** the spectral staircase function $N(E)$. The dashed lines denote the mean density of states $\bar{d}(E)$ and the average number of states $\bar{N}(E)$ discussed in more detail in sect. 33.1.1.

in (30.17) is often divergent. We shall return to this point in sects. 33.1.1 and 33.1.2.

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 30.1 (a),

$$d(E) = \sum_n \delta(E - E_n). \tag{30.18}$$

Using the identity

[exercise 30.1]

$$\delta(E - E_n) = -\lim_{\epsilon\to+0} \frac{1}{\pi} \mathrm{Im} \frac{1}{\mathrm{E} - \mathrm{E_n} + \mathrm{i}\epsilon} \tag{30.19}$$

we can express the density of states in terms of the trace of the Green's function, that is

$$d(E) = \sum_n \delta(E - E_n) = -\lim_{\epsilon\to 0} \frac{1}{\pi} \mathrm{Im}\, \mathrm{tr}\, \mathrm{G}(\mathrm{q}, \mathrm{q}', \mathrm{E} + \mathrm{i}\epsilon). \tag{30.20}$$

[section 33.1.1]

As we shall see after "some" work, a semiclassical formula for right hand side of this relation will yield the quantum spectrum in terms of periodic orbits.

The density of states can be written as the derivative $d(E) = \mathrm{d}N(E)/\mathrm{d}E$ of the *spectral staircase* function

$$N(E) = \sum_n \Theta(E - E_n) \tag{30.21}$$

which counts the number of eigenenergies below $E$, figure 30.1 (b). Here $\Theta$ is the Heaviside function

$$\Theta(x) = 1 \quad \text{if } x > 0; \quad \Theta(x) = 0 \quad \text{if } x < 0\,. \tag{30.22}$$

The spectral staircase is a useful quantity in many contexts, both experimental and theoretical. This completes our lightning review of quantum mechanics.

## Exercises

30.1. **Dirac delta function, Lorentzian representation.** Derive the representation (30.19)

$$\delta(E - E_n) = -\lim_{\epsilon \to +0} \frac{1}{\pi} \mathrm{Im} \frac{1}{\mathrm{E} - \mathrm{E_n} + \mathrm{i}\epsilon}$$

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).

30.2. **Green's function.** Verify Green's function Laplace transform (30.16),

$$\begin{aligned} G(q, q', E + i\varepsilon) &= \frac{1}{i\hbar} \int_0^\infty dt\, e^{\frac{i}{\hbar}Et - \frac{\varepsilon}{\hbar}t} K(q, q', t) \\ &= \sum \frac{\phi_n(q)\phi_n^*(q')}{E - E_n + i\varepsilon} \end{aligned}$$

argue that positive $\epsilon$ is needed (hint: read a good quantum mechanics textbook).

# Chapter 31

# WKB quantization

THE WAVE FUNCTION for a particle of energy $E$ moving in a constant potential $V$ is

$$\psi = Ae^{\frac{i}{\hbar}pq} \tag{31.1}$$

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 classical 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 - classical mechanics computations yield suprisingly 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 classical flow that plays no role in classical mechanics.

## 31.1 WKB ansatz

Consider a time-independent Schrödinger equation in 1 spatial dimension:

$$-\frac{\hbar^2}{2m}\psi''(q) + V(q)\psi(q) = E\psi(q)\,, \tag{31.2}$$

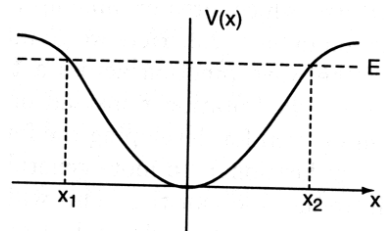



**Figure 31.1:** A 1-dimensional potential, location of the two turning points at fixed energy $E$.

with potential $V(q)$ growing sufficiently fast as $q \to \pm\infty$ so that the classical particle motion is confined for any $E$. Define the local momentum $p(q)$ and the local wavenumber $k(q)$ by

$$p(q) = \pm\sqrt{2m(E - V(q))}, \quad p(q) = \hbar k(q)\,. \tag{31.3}$$

The variable wavenumber form of the Schrödinger equation

$$\psi'' + k^2(q)\psi = 0 \tag{31.4}$$

sugests 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:

$$(S')^2 = p^2 + \hbar^2\frac{A''}{A} \tag{31.5}$$

$$S''A + 2S'A' = \frac{1}{A}\frac{d}{dq}(S'A^2) = 0\,. \tag{31.6}$$

The Wentzel-Kramers-Brillouin *(WKB)* or *semiclassical* approximation consists of dropping the $\hbar^2$ term in (31.5). 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 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 (31.5) we obtain the phase increment of a wave function initially at $q$, at energy $E$

$$S(q, q', E) = \int_{q'}^{q} dq''\, p(q'')\,. \tag{31.7}$$

This integral over a particle trajectory of constant energy, called the *action*, will play a key role in all that follows. The integration of (31.6) is even easier

$$A(q) = \frac{C}{|p(q)|^{\frac{1}{2}}}, \qquad C = |p(q')|^{\frac{1}{2}}\psi(q')\,, \tag{31.8}$$

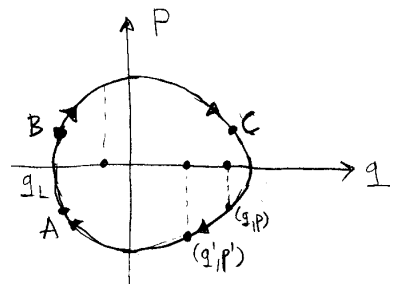



**Figure 31.2:** A 1-dof phase space trajectory of a particle moving in a bound potential.

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

$$\psi_{sc}(q, q', E) = \frac{C}{|p(q)|^{\frac{1}{2}}} e^{\frac{i}{\hbar} S(q,q',E)} \,. \tag{31.9}$$

In what follows we shall suppress dependence on the initial point and energy in such formulas, $(q, q', E) \to (q)$.

The WKB ansatz generalizes the free motion wave function (31.1), 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 31.1, where all energy is potential, and

$$p(q) \to 0 \quad \text{as} \quad q \to q_0 \,, \tag{31.10}$$

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 31.1), 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, 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's idea was to proceed from the initial point $(q', p')$ to a point $(q_A, p_A)$ preceeding the turning point in the $\psi(q)$ representation, then switch to the momentum representation

$$\widetilde{\psi}(p) = \frac{1}{\sqrt{2\pi\hbar}} \int dq\, e^{-\frac{i}{\hbar} qp} \psi(q) \,, \tag{31.11}$$

continue from $(q_A, p_A)$ to $(q_B, p_B)$, switch back to the coordinate representation,

$$\psi(q) = \frac{1}{\sqrt{2\pi\hbar}} \int dp\, e^{\frac{i}{\hbar} qp} \widetilde{\psi}(p) \,, \tag{31.12}$$

and so on.

The only rub is that one usually cannot evaluate these transforms exactly. But, as the WKB wave function (31.9) is approximate anyway, it suffices to estimate these transforms to leading order in $\hbar$ accuracy. This is accomplished by the method of stationary phase.

## 31.2 Method of stationary phase

All "semiclassical" approximations are based on saddle point evaluations of integrals of the type

$$I = \int dx\, A(x)\, e^{is\Phi(x)}\,, \qquad x, \Phi(x) \in \mathbb{R}\,, \tag{31.13}$$

where $s$ is assumed to be a large, 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 $\delta x$,

$$I = \int dx\, A(x)\, e^{is(\Phi(x_0)+\frac{1}{2}\Phi''(x_0)\delta x^2+\ldots)}\,. \tag{31.14}$$

Assume (for time being) that $\Phi''(x_0) \neq 0$, with either sign, $\mathrm{sgn}[\Phi''] = \Phi''/|\Phi''| = \pm 1$. 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

$$I \approx A(x_0) e^{is\Phi(x_0)} \int dx\, e^{\frac{1}{2}is\Phi''(x_0)(x-x_0)^2}\,. \tag{31.15}$$

Using the *Fresnel integral formula*

[exercise 31.1]

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\, e^{-\frac{x^2}{2ia}} = \sqrt{ia} = |a|^{1/2}\, e^{i\frac{\pi}{4}\frac{a}{|a|}} \tag{31.16}$$

we obtain

$$I \approx A(x_0) \left| \frac{2\pi}{s\Phi''(x_0)} \right|^{1/2} e^{is\Phi(x_0)\pm i\frac{\pi}{4}}\,, \tag{31.17}$$

where $\pm$ corresponds to the positive/negative sign of $s\Phi'(x_0)$.

## 31.3 WKB quantization

We can now evaluate the Fourier transforms (31.11), (31.12) to the same order in $\hbar$ as the WKB wave function using the stationary phase method,

$$\begin{aligned}\widetilde{\psi}_{sc}(p) &= \frac{C}{\sqrt{2\pi\hbar}}\int\frac{dq}{|p(q)|^{\frac{1}{2}}}\,e^{\frac{i}{\hbar}(S(q)-qp)}\\ &\approx \frac{C}{\sqrt{2\pi\hbar}}\frac{e^{\frac{i}{\hbar}(S(q^*)-q^*p)}}{|p(q^*)|^{\frac{1}{2}}}\int dq\,e^{\frac{i}{2\hbar}S''(q^*)(q-q^*)^2}\,,\end{aligned}\tag{31.18}$$

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 (31.16)

$$\widetilde{\psi}_{sc}(p) = \frac{C}{|p(q^*)p'(q^*)|^{\frac{1}{2}}}e^{\frac{i}{\hbar}[S(q^*)-q^*p]+\frac{i\pi}{4}\mathrm{sgn}[S''(q^*)]}\,.\tag{31.19}$$

As we continue from $(q_A, p_A)$ to $(q_B, p_B)$, nothing problematic occurrs - $p(q^*)$ 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 expresses everything in terms of the $p$ variable,

$$q^* = q(p), \quad \frac{d}{dq}q = 1 = \frac{dp}{dq}\frac{dq(p)}{dp} = q'(p)p'(q^*)\,.\tag{31.20}$$

As the classical trajectory crosses $q_L$, the weight in (31.19),

$$\frac{d}{dq}p^2(q_L) = 2p(q_L)p'(q_L) = -2mV'(q)\,,\tag{31.21}$$

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 (31.19) is $-\frac{\pi}{4}$. To go back from the $p$ to the $q$ representation, just turn figure 31.2 $90^o$ anticlockwise. Everything is the same if you replace $(q, p) \to (-p, q)$; so, without much ado we get the semiclassical wave function at the point $(q_B, p_B)$,

$$\psi_{sc}(q) = \frac{e^{\frac{i}{\hbar}(\tilde{S}(p^*)+qp^*)-\frac{i\pi}{4}}}{|q^*(p^*)|^{\frac{1}{2}}}\widetilde{\psi}_{sc}(p^*) = \frac{C}{|p(q)|^{\frac{1}{2}}}e^{\frac{i}{\hbar}S(q)-\frac{i\pi}{2}}\,.\tag{31.22}$$

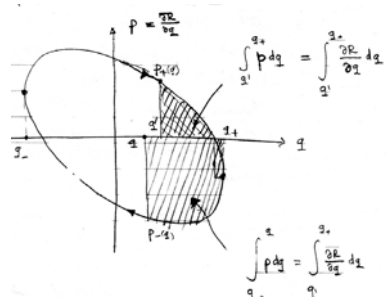


**Figure 31.3:** $S_p(E)$, the action of a periodic orbit $p$ at energy $E$, equals the area in the phase space traced out by the 1-dof trajectory.

The extra $|p'(q^*)|^{1/2}$ weight in (31.19) is cancelled by the $|q'(p^*)|^{1/2}$ term, by the Legendre relation (31.20).

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 31.2 by $180^o$, and flipping coordinates $(q, p) \to (-q, -p)$. While a turning point is not an invariant concept (for a sufficiently short trajectory segment, it can be undone by a $45^o$ 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 (31.8), we obtain

$$\psi(q') = \psi(q) = \left|\frac{p(q')}{p(q)}\right|^{\frac{1}{2}} e^{i(\frac{1}{\hbar}\oint p(q)dq - \pi)}\psi(q') \,.$$

The prefactor is 1 by the periodic orbit condition $q = q'$, so the phase must be a multiple of $2\pi$,

$$\frac{1}{\hbar}\oint p(q)dq = 2\pi\left(n + \frac{m}{4}\right) , \tag{31.23}$$

where $m$ is the number of turning points along the trajectory - for this 1-dof problem, $m = 2$.

The action integral in (31.23) is the area (see figure 31.3) enclosed by the classical phase space loop of figure 31.2, and the quantization condition says that eigenenergies 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 most recent quantum chaotic (re)birth until the 1970's. Despite its derivation in a noninvariant coordinate frame, the final result involves only canonically invariant classical quantities, the periodic orbit action $S$, and the topological index $m$.

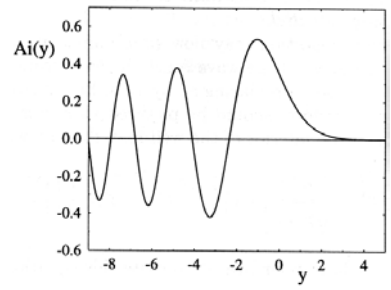



**Figure 31.4:** Airy function $Ai(q)$.

### 31.3.1 Harmonic oscillator quantization

Let us check the WKB quantization for one case (the only case?) whose quantum mechanics we fully understand: the harmonic oscillator

$$E = \frac{1}{2m}\left(p^2 + (m\omega q)^2\right) .$$

The loop in figure 31.2 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

$$E_n = \hbar\omega(n + 1/2) \tag{31.24}$$

turns out to be the *exact* harmonic oscillator spectrum. The stationary phase condition (31.18) 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 (31.5) as a short wavelength approximation.

## 31.4 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 (31.16), 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 (31.14) fails whenever $\Phi''(x) = 0$, or, in our the WKB ansatz (31.18), whenever the momentum $p'(q) = S''(q)$ vanishes. In that case we have to go beyond the quadratic approximation (31.15) to the first nonvanishing term in the Taylor expansion of the exponent. If $\Phi'''(x_0) \neq 0$, then

$$I \approx A(x_0)e^{is\Phi(x_0)} \int_{-\infty}^{\infty} dx\, e^{is\Phi'''(x_0)\frac{(x-x_0)^3}{6}} . \tag{31.25}$$

Airy functions can be represented by integrals of the form

$$Ai(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} dy\, e^{i(xy-\frac{y^3}{3})}\,. \tag{31.26}$$

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

$$\psi'' = q\psi\,, \tag{31.27}$$

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 and learning about their asymptotics - a challenge that we will have to eventually overcome, in order to incorporate diffraction phenomena into semiclassical quantization.

2) what does the wave function look like?

3) generically useful when Gaussian approximations fail

The physical origin of the topological phase is illustrated by the shape of the Airy function, figure 31.4. For a potential with a finite slope $V'(q)$ the wave function pentrates into the forbidden region, and accomodates a bit more of a stationary wavelength then what one would expect from the classical 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 classical trajectory. While in the $q$-representation the WKB ansatz a turning point is singular, along the $p$ direction the classical 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 classical 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 classical trajectory in the $p$-space until you encounter the next $p$-space turning point; go back to the $q$-space representation, an 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 semiclassical 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 $a$, $V(q) = V(a) + (q-a)V'(a) + \cdots$, 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 inbetween via the WKB connection formulas. The single-valuedness condition again yields the 1-dimensional WKB quantization. This a bit more work than tracking the classical 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 31.1** Airy function. The stationary phase approximation is all that is needed for the semiclassical approximation, with the proviso that $D$ in (32.36) 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 [10].

[exercise 31.4]

**Remark 31.2** Bohr-Sommerfeld quantization. Bohr-Sommerfeld quantization condition was the key result of the old quantum theory, in which the electron trajectories were purely classical. 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 [5] added the topological indices.

## Exercises

31.1. **WKB ansatz.** Try to show that no other ansatz other than (32.1) gives a meaningful definition of the momentum in the $\hbar \to 0$ limit.

31.2. **Fresnel integral.** Derive the Fresnel integral

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\, e^{-\frac{x^2}{2ia}} = \sqrt{ia} = |a|^{1/2} e^{i\frac{\pi}{4}\frac{a}{|a|}} .$$

31.3. **Sterling formula for $n!$.** Compute an approximate value of $n!$ for large $n$ using the stationary phase approx-

imation. Hint: $n! = \int_0^\infty dt\, t^n e^{-t}$.

31.4. **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

$$Ai(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dy\, e^{i(xy - \frac{y^3}{3})} . \tag{31.28}$$

Calculate the Airy function $Ai(x)$ using the stationary phase approximation. What happens when considering the limit $x \to 0$. Estimate for which value of $x$ the stationary phase approximation breaks down.

## References


[31.1] D. J. Griffiths, *Introduction to Quantum Mechanics* (Prentice-Hall, Englewood Cliffs, New Jersey, 1994).

[31.2] J.W.S. Rayleigh, *The Theory of Sound* (Macmillan, London 1896; reprinted by Dover, New York 1945).

[31.3] J.B. Keller, "Corrected Bohr-Sommerfeld quantum conditions for nonseparable systems," *Ann. Phys. (N.Y.)* **4**, 180 (1958).

[31.4] J.B. Keller and S.I. Rubinow, *Ann. Phys. (N.Y.)* **9**, 24 (1960).

[31.5] J.B. Keller, "A geometrical theory of diffraction," in *Calculus of variations and its applications*, *Proc. of Symposia in appl. math.* **8**, (McGraw-Hill, New York, 1958).

[31.6] J.B. Keller, *Calculus of Variations* **27**, (1958).

[31.7] V.P. Maslov, *Théorie des Perturbations et Métodes Asymptotiques* (Dunod, Paris, 1972).

[31.8] V.P. Maslov and M.V. Fedoriuk, *Semi-Classical Approximation in Quantum Mechanics* (Reidel, Boston 1981).

[31.9] V.I. Arnold, *Functional Anal. Appl.* **1**, 1 (1967).

[31.10] N. Bleistein and R.A. Handelsman, *Asymptotic Expansions of Integrals* (Dover, New York 1986).

[31.11] I.C. Percival, *Adv. Chem. Phys.* **36**, 1 (1977).

# Chapter 32

# 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 classical mechanics. The novelty of quantum mechanics is the interference of the point-like particle with other versions of itself traveling along different classical trajectories, a feat impossible in classical mechanics. The short wavelength – or semiclassical – formalism is developed by formally taking the limit $\hbar \to 0$ in quantum mechanics in such a way that quantum quantities go to their classical counterparts. [remark 32.1]

## 32.1 Hamilton-Jacobi theory

We saw in chapter 31 that for a 1-dof particle moving in a slowly varying potential, it makes sense to generalize the free particle wave function (31.1) to a wave function

$$\psi(q,t) = A(q,t)e^{iR(q,t)/\hbar}\,, \tag{32.1}$$

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 [exercise 31.1]

$\psi$ follows from the Schrödinger equation (30.1)

$$\left(i\hbar\frac{\partial}{\partial t} + \frac{\hbar^2}{2m}\frac{\partial^2}{\partial q^2} - V(q)\right)\psi(q,t) = 0\,. \tag{32.2}$$

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

$$\frac{\partial R}{\partial t} + \frac{1}{2m}\left(\frac{\partial R}{\partial q}\right)^2 + V(q) - \frac{\hbar^2}{2m}\frac{1}{A}\frac{\partial^2}{\partial q^2}A = 0\,, \tag{32.3}$$

and the imaginary part the time evolution of the amplitude

[exercise 32.6]
[exercise 32.7]

$$\frac{\partial A}{\partial t} + \frac{1}{m}\sum_{i=1}^{D}\frac{\partial A}{\partial q_i}\frac{\partial R}{\partial q_i} + \frac{1}{2m}A\frac{\partial^2 R}{\partial q^2} = 0\,. \tag{32.4}$$

[exercise 32.8]

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 (32.3) is, however, of order $\hbar^2$ and thus small in the semiclassical limit $\hbar \to 0$.

Now we generalize the *Wentzel-Kramers-Brillouin* (WKB) *ansatz* for 1-dof 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

$$\frac{\partial R}{\partial t} + H\left(q, \frac{\partial R}{\partial q}\right) = 0\,. \tag{32.5}$$

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.

### 32.1.1 Hamilton's equations

We now solve the nonlinear partial differential equation (32.5) in a way the 17 year old Hamilton might have solved it. The main step is the step leading from the nonlinear PDE (32.9) to Hamilton's ODEs (32.10). If you already understand the Hamilton-Jacobi theory, you can safely skip this section.

fast track:
sect. 32.1.3, p. 527

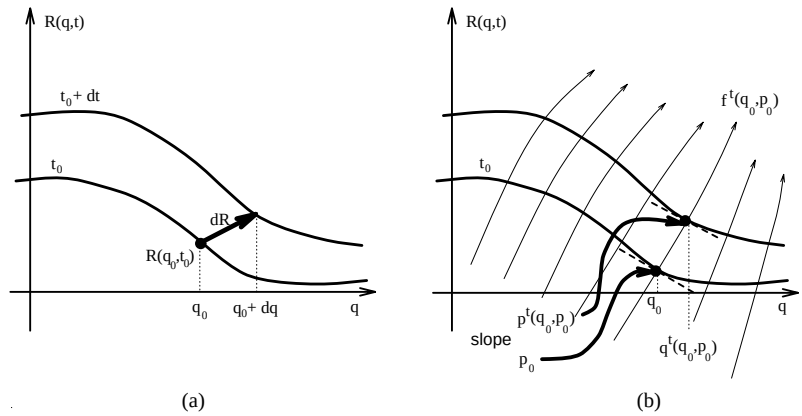



**Figure 32.1:** (a) A phase $R(q,t)$ plotted as a function of the position $q$ for two infinitesimally close times. (b) The phase $R(q,t)$ transported by a swarm of "particles"; The Hamilton's equations (32.10) construct $R(q,t)$ by transporting $q_0 \to q(t)$ and the slope of $R(q_0,t_0)$, that is $p_0 \to p(t)$.

The wave equation (30.1) describes how the wave function $\psi$ evolves with time, and if you think of $\psi$ as an (infinite dimensional) vector, position $q$ plays a 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 32.1 (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 32.1 (b). For notational convenience, define

$$p_i = p_i(q,t) := \frac{\partial R}{\partial q_i}\,, \quad i = 1,2,\ldots,D\,. \tag{32.6}$$

We saw earlier that (32.3) reduces in the semiclassical approximation to the Hamilton-Jacobi equation (32.5). 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 32.1 (a)

$$dR = \frac{\partial R}{\partial t}dt + \frac{\partial R}{\partial q}dq\,. \tag{32.7}$$

Dividing through by $dt$ and substituting (32.5) we obtain the total derivative of $R(q,t)$ with respect to time *along the as yet arbitrary direction* $\dot{q}$, that is,

$$\frac{dR}{dt}(q,\dot{q},t) = -H(q,p) + \dot{q}\cdot p\,. \tag{32.8}$$

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 (32.8) 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 (32.5) 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,

$$\dot{p}(q,\dot{q},t) + \frac{\partial H}{\partial q} = \left(\dot{q} - \frac{\partial H}{\partial p}\right)\frac{\partial p}{\partial q}\,. \tag{32.9}$$

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

$$\dot{q} = \frac{\partial H(q,p)}{\partial p}, \qquad \dot{p} = -\frac{\partial H(q,p)}{\partial q} \tag{32.10}$$

with initial conditions

$$q(t_0) = q', \qquad p(t_0) = p' = \frac{\partial R}{\partial q}(q', t_0). \tag{32.11}$$

[section 7.1]

We recognize (32.10) as Hamilton's equations of motion of classical mechanics. The miracle happens in the step leading from (32.5) to (32.9) – 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.

$\dot{q}$ is no longer an independent function, and the phase $R(q,t)$ can now be computed by integrating equation (32.8) along the trajectory $(q(t), p(t))$

$$\begin{aligned} R(q,t) &= R(q',t_0) + R(q,t;q',t_0) \\ R(q,t;q',t_0) &= \int_{t_0}^{t} d\tau\left[\dot{q}(\tau)\cdot p(\tau) - H(q(\tau),p(\tau))\right]\,, \end{aligned} \tag{32.12}$$

with the initial conditions (32.11). In this way the Hamilton-Jacobi *partial* differential equation (32.3) 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 32.1 (b). The integrand of (32.12) is known as the *Lagrangian*,

$$L(q, \dot{q}, t) = \dot{q} \cdot p - H(q, p, t) \,. \tag{32.13}$$

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 $q$:

$$p' = \frac{\partial}{\partial q'} R(q', 0) = -\frac{\partial}{\partial q'} R(q, q', t). \tag{32.14}$$

[exercise 32.5]

The function $R(q, q', t)$ is known as *Hamilton's principal function*.

[exercise 32.9]

To summarize: Hamilton's achievement was to trade in the Hamilton-Jacobi *partial* differential equation (32.5) 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 (32.12) evaluated along the phase space trajectory $(q(\tau), p(\tau))$.

### 32.1.2 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 (32.12) is simply $Et$. The first term, or the *action*

$$S(q, q', E) = \int_0^t d\tau \, \dot{q}(\tau) \cdot p(\tau) = \int_{q'}^{q} dq \cdot p \tag{32.15}$$

is integrated along a trajectory from $q'$ to $q$ with a fixed energy $E$. By (32.12) the action is a Legendre transform of Hamilton's principal function

$$S(q, q', E) = R(q, q', t) + Et \,. \tag{32.16}$$

The time of flight $t$ along the trajectory connecting $q' \to q$ with fixed energy $E$ is given by

$$\frac{\partial}{\partial E} S(q, q', E) = t\,. \tag{32.17}$$

The way to think about the formula (32.16) 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 $q' \to q$ with a given time of flight $t$ is given by the derivative of Hamilton's principal function

$$\frac{\partial}{\partial t} R(q, q', t) = -E\,, \tag{32.18}$$

and the second variations of $R$ and $S$ are related in the standard way of Legendre transforms:

$$\frac{\partial^2}{\partial t^2} R(q, q', t) \frac{\partial^2}{\partial E^2} S(q, q', E) = -1\,. \tag{32.19}$$

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 $\gamma$ in it has zero action, [section 32.1.4]

$$0 = \oint_\gamma dq \cdot p,$$

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.

### 32.1.3 Density evolution

To obtain the full solution of the Schrödinger equation (30.1), we also have to integrate (32.4).

$$\rho(q, t) := A^2 = \psi^* \psi$$

plays the role of a density. To the leding 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 (32.4) is equivalent to the continuity equation (14.36) after multiplying (32.4) by $2A$, that is

$$\frac{\partial\rho}{\partial t} + \frac{\partial}{\partial q_i}(\rho v_i) = 0\,. \tag{32.20}$$

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 (32.5), 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 $(q(0) = q', p(0) = p' = \partial_q R(q',0))$.

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 $\rho(q,t)d^D dq$. 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 32.2. The ratio of the initial and the final volumes can be expressed as

$$\rho(q(t),t) = \left|\det\frac{\partial q'}{\partial q}\right|\rho(q',0)\,. \tag{32.21}$$

[section 14.2]

As we know how to compute trajectories $(q(t), p(t))$, we know how to compute this Jacobian and, by (32.21), the density $\rho(q(t),t)$ at time $t$.

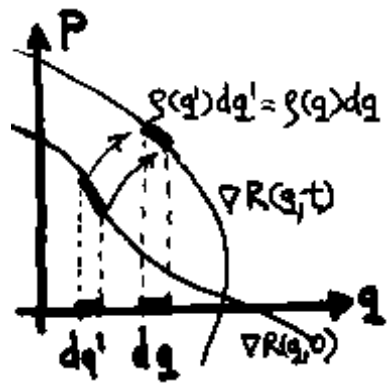



**Figure 32.2:** Density evolution of an initial surface $(q', p' = \partial_q R(q', 0)$ into $(q(t), p(t))$ surface time $t$ later, sketched in 1 dimension. While the number of trajectories and the phase space Liouville volume are conserved, the density of trajectories projected on the $q$ coordinate varies; trajectories which started in $dq'$ at time zero end up in the interval $dq$.

### 32.1.4 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 32.3.

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 32.3, 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 trough 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

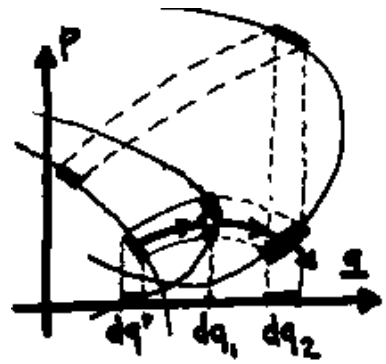



**Figure 32.3:** Folding of the Lagrangian surface $(q, \partial_q R(q,t))$.

$q'$ 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

$$\det \left.\frac{\partial q'}{\partial q}\right|_j = e^{-i\pi m_j(q,q',t)} \left|\det \left.\frac{\partial q'}{\partial q}\right|_j\right| ,$$

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 32.3. 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 inital $q'$ and end in $q$ in time $t$

$$\psi_{sc}(q,t) = \int dq' \sum_j \left|\det \left.\frac{\partial q'}{\partial q}\right|_j\right|^{1/2} e^{iR_j(q,q',t)/\hbar - i\pi m_j(q,q',t)/2} \psi(q'_j, 0) , \qquad (32.22)$$

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.

## 32.2 Semiclassical propagator

We saw in chapter 30 that the evolution of an initial wave function $\psi(q,0)$ is completely determined by the propagator (30.12). As $K(q,q',t)$ itself satisfies the

Schrödinger equation (30.14), 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 (30.15) demands that the propagator at $t = 0$ is a $\delta$-function at $q = q'$, that is, the amplitude is infinite at $q'$ 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 at a semiclassical expression for $K(q, q', t)$ by considering the propagator for short times first, and extrapolating from there to arbitrary times $t$.

### 32.2.1 Short time propagator

For infinitesimally short times $\delta t$ 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 (32.12), that is

$$R(q, q', \delta t) = p\dot{q}\delta t - H(q, p)\delta t , \tag{32.23}$$

with $\dot{q} \approx (q - q')/\delta t$. For Hamiltonians of the form (30.2) 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

$$K_{sc}(q, q', \delta t) \approx A(q, q', \delta t) e^{\frac{i}{\hbar}\left(\frac{m}{2\delta t}(q-q')^2 - V(q)\delta t\right)} . \tag{32.24}$$

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

$$\delta(z) = \lim_{\sigma \to 0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2\sigma^2} \tag{32.25}$$

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 $\delta t$ [exercise 32.1]

is therefore

$$K_{sc}(q,q',\delta t) \approx \left(\frac{m}{2\pi i\hbar\delta t}\right)^{D/2} e^{\frac{i}{\hbar}\left(\frac{m(q-q')^2}{2\delta t}-V(q)\delta t\right)} . \tag{32.26}$$

The short time dynamics of the Lagrangian manifold $(q, \partial_q R)$ which corresponds to the quantum propagator can now be deduced from (32.23); one obtains

$$\frac{\partial R}{\partial q} = p \approx \frac{m}{\delta t}(q-q') ,$$

i.e., is the particles start for short times on a Lagrangian manifold which is a plane in phase space, see figure 32.4. Note, that for $\delta t \to 0$, 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 (32.26) 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

$$K_{sc}(q,q',\delta t) = \frac{1}{(2\pi i\hbar)^{D/2}}\left(\det\frac{\partial p'}{\partial q}\right)^{1/2} e^{iR(q,q',\delta t)/\hbar}, \tag{32.27}$$

where

$$\left.\frac{\partial p'_i}{\partial q_j}\right|_{t,q'} = \frac{\partial^2 R(q,q',\delta t)}{\partial q_j \partial q'_i} \tag{32.28}$$

The subscript $\cdots|_{t,q'}$ indicates that the partial derivatives are to be evaluated with $t, q'$ fixed.

The propagator in (32.27) 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 (32.22)

$$K_{sc}(q'',q',t'+\delta t) = \sum_j \left|\det\frac{\partial q}{\partial q''}\right|_j^{1/2} e^{iR_j(q'',q,t')/\hbar - i\pi m_j(q'',q,t')/2} K(q,q'_j,\delta t) ,$$

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 $m_j =$

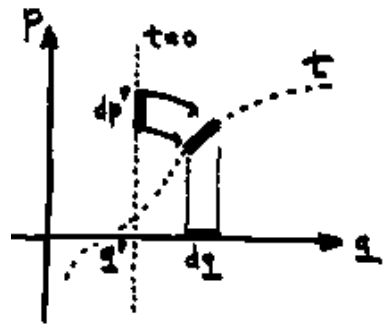



**Figure 32.4:** Evolution of the semiclassical propagator. The configuration which corresponds to the initial conditions of the propagator is a Lagrangian manifold $q = q'$, that is, a plane parallel to the $p$ axis. The hypothetical particles are thus initially all placed at $q'$ but take on all possible momenta $p'$. The Jacobian matrix $C$ (32.29) 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

$$\det \left.\frac{\partial q}{\partial q''}\right|_t \det \left.\frac{\partial p'}{\partial q}\right|_{q',\delta t} = \det \left.\frac{\partial p'}{\partial q''}\right|_{q',t'+\delta t} . \tag{32.29}$$

The final form of the semiclassical or *Van Vleck propagator*, is thus

$$K_{sc}(q, q', t) = \sum_j \frac{1}{(2\pi i\hbar)^{D/2}} \left|\det \frac{\partial p'}{\partial q}\right|^{1/2} e^{iR_j(q,q',t)/\hbar - i m_j \pi/2} . \tag{32.30}$$

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 (32.30) with some care. Unlike the full quantum propagator, which satisfies the group property (30.13) exactly, the semiclassical propagator performs this only approximately, that is

$$K_{sc}(q, q', t_1 + t_2) \approx \int dq''\, K_{sc}(q, q'', t_2) K_{sc}(q'', q', t_1) . \tag{32.31}$$

The connection can be made explicit by the stationary phase approximation, sect.31.2. Approximating the integral in (32.31) by integrating only over regions near points $q''$ at which the phase is stationary, leads to the stationary phase condition

$$\frac{\partial R(q, q'', t_2)}{\partial q''_i} + \frac{\partial R(q'', q', t_1)}{\partial q''_i} = 0. \tag{32.32}$$

Classical 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

classical evolution of sect. 15.2, 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 classical momenta at each evolution step.

### 32.2.2 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 (32.12) is

$$R(q,q',t) = \frac{m(q-q')^2}{2t}\,. \tag{32.33}$$

The weight $\det\frac{\partial p'}{\partial q}$ from (32.28) can be evaluated explicitly, and the Van Vleck propagator is

$$K_{sc}(q,q',t) = \left(\frac{m}{2\pi i\hbar t}\right)^{D/2} e^{im(q-q')^2/2\hbar t}\,, \tag{32.34}$$

identical to the short time propagator (32.26), 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 (32.2). 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 classical) paths from $q'$ to $q$. [remark 32.3]

[exercise 32.10]
[exercise 32.11]
[exercise 32.12]

## 32.3 Semiclassical Green's function

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 (30.1). Even though we assumed in the calculation a time independent Hamiltonian of the special form (30.2), the derivation would lead to the same final result (32.30) 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 (30.10). It is therefore advantageous to switch from a time representation to an energy representation, that is from the propagator (30.12) to the energy dependent Green's function (30.16). A semiclassical approximation of the Green's function $G_{sc}(q,q',E)$ is given by the Laplace transform (30.16) of the Van Vleck propagator $K_{sc}(q,q',t)$:

$$G_{sc}(q,q',E) = \frac{1}{i\hbar}\int_0^\infty dt\, e^{iEt/\hbar} K_{sc}(q,q',t)\,. \tag{32.35}$$

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.

### 32.3.1 Stationary phase in higher dimensions

[exercise 31.1]

Generalizing the method of sect. 31.2 to $d$ dimensions, consider stationary phase points fulfilling

$$\frac{d}{dx_i}\Phi(x)\bigg|_{x=x_0} = 0 \quad \forall i = 1, \dots d\,.$$

An expansion of the phase up to second order involves now the symmetric matrix of second derivatives of $\Phi(x)$, that is

$$D_{ij}(x_0) = \frac{\partial^2}{\partial x_i \partial x_j}\Phi(x)\bigg|_{x=x_0}\,.$$

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 (31.13) is then

$$I \approx \sum_{x_0} (2\pi i/s)^{d/2}\, |\det D(x_0)|^{-1/2} A(x_0)\, e^{is\Phi(x_0) - \frac{i\pi}{2} m(x_0)}\,, \tag{32.36}$$

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 26.2]
[exercise 32.2]
[exercise 31.3]

The stationary phase approximation is all that is needed for the semiclassical approximation, with the proviso that $D$ in (32.36) has no zero eigenvalues.

### 32.3.2 Long trajectories

When evaluating the integral (32.35) 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 $t \to 0$ 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. 32.3.3.

The stationary phase points $t^*$ of the integrand in (32.35) are given by the condition

$$\frac{\partial}{\partial t} R(q, q', t^*) + E = 0\,. \tag{32.37}$$

We recognize this condition as the solution of (32.18), 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,

$$R(q, q', t) + Et = R(q, q', t^*) + Et^* + \frac{1}{2}(t - t^*)^2 \frac{\partial^2}{\partial t^2} R(q, q', t^*) + \cdots$$

the stationary phase approximation of the integral corresponding to a classical trajectory $j$ in the Van Vleck propagator sum (32.30) yields

$$G_j(q, q', E) = \frac{1}{i\hbar(2i\pi\hbar)^{(D-1)/2}} \left| \det C_j \left( \frac{\partial^2 R_j}{\partial t^2} \right)^{-1} \right|^{1/2} e^{\frac{i}{\hbar} S_j - \frac{i\pi}{2} m_j}, \tag{32.38}$$

where $m_j = m_j(q, q', E)$ now includes a possible additional phase arising from the time stationary phase integration (31.16), 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 (32.16)

$$S(q, q', E) = R(q, q', t^*) + Et^*, \qquad \text{with} \quad t^* = t^*(q, q', E), \tag{32.39}$$

the Legendre transform of Hamilton's principal function. Note that the partial derivative of the action (32.39) 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

$$\frac{\partial S(q, q', E)}{\partial q_i} = \frac{\partial R(q, q', t^*)}{\partial q_i}, \tag{32.40}$$

due to the stationary phase condition (32.37), 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 32.13]

Next we will simplify the amplitude term in (32.38) and rewrite it as an explicit function of the energy. Consider the $[(D+1)\times(D+1)]$ matrix

$$D(q, q', E) = \begin{pmatrix} \frac{\partial^2 S}{\partial q' \partial q} & \frac{\partial^2 S}{\partial q' \partial E} \\ \frac{\partial^2 S}{\partial q \partial E} & \frac{\partial^2 S}{\partial E^2} \end{pmatrix} = \begin{pmatrix} -\frac{\partial p'}{\partial q} & -\frac{\partial p'}{\partial E} \\ \frac{\partial t}{\partial q} & \frac{\partial t}{\partial E} \end{pmatrix}, \tag{32.41}$$

where $S = S(q, q', E)$ and we used (32.14–32.17) here to obtain the left hand side of (32.41). The minus signs follow from observing from the definition of (32.15) that $S(q, q', E) = -S(q', q, E)$. Note that $D$ is nothing but the Jacobian matrix

of the coordinate transformation $(q, E) \to (p', t)$ 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 $(\det .)_{q',t}$ for a Jacobian determinant with partial derivatives evaluated at $t, q'$ fixed, and likewise for other subscripts. Using the relation (32.19) 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 (32.28) and the amplitude factor arising from the stationary phase approximation. The amplitude in (32.38) 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 (32.28).

We can now write down the semiclassical approximation of the contribution of the $j$th trajectory to the Green's function (32.38) in explicitly energy dependent form:

$$G_j(q, q', E) = \frac{1}{i\hbar(2i\pi\hbar)^{(D-1)/2}}\left|\det D_j\right|^{1/2} e^{\frac{i}{\hbar}S_j - \frac{i\pi}{2}m_j}. \tag{32.42}$$

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

$$\begin{aligned}\frac{\partial}{\partial q'_i}H(q,p) &= 0 = \frac{\partial H}{\partial p_j}\frac{\partial p_j}{\partial q'_i} = \dot{q}_j\frac{\partial^2 S}{\partial q_j \partial q'_i}\\ \frac{\partial}{\partial q_i}H(q',p') &= 0 = \frac{\partial^2 S}{\partial q_i \partial q'_j}\dot{q}'_j\,,\end{aligned} \tag{32.43}$$

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

$$(q_1, q_2, q_3, \cdots, q_d) \to (q_\parallel, q_{\perp 1}, q_{\perp 2}, \cdots, q_{\perp(D-1)})$$

so that one axis points along the trajectory and all others are perpendicular to it

$$(\dot{q}_1, \dot{q}_2, \dot{q}_3, \cdots, \dot{q}_d) \to (\dot{q}, 0, 0, \cdots, 0)\,.$$

With such local coordinate systems at both ends, with the longitudinal coordinate axis $q_{\|}$ 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 (32.43) takes form

$$\dot{q}\frac{\partial^2 S}{\partial q_{\|}\partial q'_i} = \frac{\partial^2 S}{\partial q_i \partial q'_{\|}}\dot{q}' = 0\,.$$

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 (32.41) is of the form

$$\det D(q, q', E) = (-1)^{D+1}\left(\det \begin{array}{ccc} 0 & 0 & \frac{\partial^2 S}{\partial E \partial q'_{\|}} \\ 0 & \frac{\partial^2 S}{\partial q_\perp \partial q'_\perp} & * \\ \frac{\partial^2 S}{\partial q_{\|}\partial E} & * & * \end{array}\right). \tag{32.44}$$

The corner entries can be evaluated using (32.17)

$$\frac{\partial^2 S}{\partial q_{\|}\partial E} = \frac{\partial}{\partial q_{\|}} t = \frac{1}{\dot{q}}\,, \qquad \frac{\partial^2 S}{\partial E \partial q'_{\|}} = \frac{1}{\dot{q}'}\,.$$

As the $q_{\|}$ 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 $D_\perp(q, q', E)$

$$\begin{aligned} \det D(q, q', E) &= \frac{1}{\dot{q}\dot{q}'}\det D_\perp(q, q', E) \\ D_\perp(q, q', E)_{ik} &= -\frac{\partial^2 S(q, q', E)}{\partial q_{\perp i}\partial q'_{\perp k}}\,. \end{aligned} \tag{32.45}$$

Putting everything together we obtain the $j$th trajectory contribution to the semiclassical Green's function

[exercise 32.15]

$$G_j(q, q', E) = \frac{1}{i\hbar(2\pi i\hbar)^{(D-1)/2}}\frac{1}{|\dot{q}\dot{q}'|^{1/2}}\left|\det D_\perp^j\right|^{1/2} e^{\frac{i}{\hbar}S_j - \frac{i\pi}{2}m_j}, \tag{32.46}$$

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 $j$.

### 32.3.3 Short trajectories

The stationary phase method cannot be used when $t^*$ is small, both because we cannot extend the integration in (31.16) 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 (32.26)

$$G_0(q, q', E) = \frac{1}{i\hbar} \int_0^\infty dt \left(\frac{m}{2\pi i\hbar t}\right)^{D/2} e^{\frac{i}{\hbar}\left(\frac{m(q-q')^2}{2t} - V(q)t + Et\right)} . \tag{32.47}$$

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

$$G_0(q, q', E) \approx -\frac{im}{2\hbar^2} \left(\frac{\sqrt{2m(E - V)}}{2\pi\hbar|q - q'|}\right)^{\frac{D-2}{2}} H^+_{\frac{D-2}{2}}(S_0(q, q', E)/\hbar) . \tag{32.48}$$

Hankel functions are stabdard, and their the 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 classical 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

classical 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 classical trajectory that connects $q' \to q$ 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*

$$\begin{aligned} G_{sc}(q,q',E) &= G_0(q,q',E) + \sum_j G_j(q,q',E) \\ G_j(q,q',E) &= \frac{1}{i\hbar(2\pi i\hbar)^{\frac{(D-1)}{2}}} \left|\frac{1}{\dot{q}\dot{q}'}\det \frac{\partial p'_\perp}{\partial q_\perp}\right|_j^{1/2} e^{\frac{i}{\hbar}S_j - \frac{i\pi}{2}m_j} \end{aligned} \tag{32.49}$$

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 (32.46) going from $q'$ to $q$ with fixed energy $E$, with $S_j(q,q',E) \gg \hbar$.

## Commentary

**Remark 32.1** Limit $\hbar \to 0$. The semiclassical limit "$\hbar \to 0$" discussed in sect. 32.1 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 [8] is 1.054571596(82) $10^{-34}$ Js.

**Remark 32.2** Madelung's fluid dynamics. Already Schrödinger [3] 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 (32.3–32.4) has been given by Madelung [2], and then built upon by Bohm [6] and others [3, 7]. Keeping the $\hbar$ dependent term in (32.3), the ordinary differential equations driving the flow (32.10) have to be altered; if the Hamiltonian can be written as kinetic plus potential term $V(q)$ as in (30.2), the $\hbar^2$ term modifies the $p$ equation of motion as

$$\dot{p}_i = -\frac{\partial}{\partial q_i}\left(V(q) + Q(q,t)\right) , \tag{32.50}$$

where, for the example at hand,

$$Q(q,t) = -\frac{\hbar^2}{2m}\frac{1}{\sqrt{\rho}}\frac{\partial^2}{\partial q^2}\sqrt{\rho} \tag{32.51}$$

interpreted by Bohm [6] as the "quantum potential." Madelung observed that Hamilton's equation for the momentum (32.50) can be rewritten as

$$\frac{\partial v_i}{\partial t} + \left(v \cdot \frac{\partial}{\partial q}\right) v_i = -\frac{1}{m}\frac{\partial V}{\partial q_i} - \frac{1}{m\rho}\frac{\partial}{\partial q_j}\sigma_{ij} , \tag{32.52}$$

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 [3] in sect. 32.1.3. 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 classical dynamics corresponding to quantum evolution is thus that of an "hypothetical fluid" experiencing $\hbar$ and $\rho$ dependent stresses. The "hydrodynamic" interpretation of quantum mechanics has, however, not been very fruitful in practice.

**Remark 32.3** Path integrals. The semiclassical propagator (32.30) 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 (32.34) 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 classical) 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 [4] added the topological indices when deriving the semiclassical propagator from Feynman's path integral by stationary phase conditions.

**Remark 32.4** 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 32.5** The quasiclassical approximation The *quasiclassical* approximation was introduced by Maslov[**?**]. The term 'quasiclassical' is more appropriate than semiclassical since the Maslov type description leads to a pure classical evolution operator in a natural way. Following mostly ref. [**?**], we give a summary of the quasiclassical approximation, which was worked out by Maslov[**?**] in this form. One additional advantage of this description is that the wave function evolves along one single classical trajectory and we do not have to compute sums over increasing numbers of classical trajectories as in computations involving Van Vleck formula[27].

## Exercises

32.1. **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 $\sigma \to 0$ limit this is the Dirac delta function

$$\int_{\mathcal{M}} dx\, \delta(x) = 1 \text{ if } 0 \in \mathcal{M}, \text{ zero otherwise} .$$

32.2. **Stationary phase approximation in higher dimensions.** All semiclassical approximations are based on saddle point evaluations of integrals of type

$$I = \int d^D x A(x) e^{i\Phi(x)/\hbar} \tag{32.53}$$

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.

32.3. **Schrödinger equation in the Madelung form.** Verify the decomposition of Schrödinger equation into real and imaginary parts, eqs. (32.3) and (32.4).

32.4. **Transport equations.** Write the wavefunction in the asymptotic form

$$\psi(q,t) = e^{\frac{i}{\hbar}R(x,t)+\frac{i}{\hbar}\varepsilon t} \sum_{n\geq 0} (i\hbar)^n A_n(x,t) .$$

Derive the transport equations for the $A_n$ by substituting this into the Schrödinger equation and then collecting terms by orders of $\hbar$. Notice that equation for $\dot{A}_n$ only requires knowledge of $A_{n-1}$ and $R$.

32.5. **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 32.13.)

32.6. **1-dimensional harmonic oscillator.** Take a 1-dimensional harmonic oscillator $U(q) = \frac{1}{2}kq^2$. 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 (32.3) and (32.4) and solve them. (Continuation: exercise 32.9.)

32.7. **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 (32.3) and (32.4).

32.8. **$D$-dimensional quadratic potentials.** Generalize the above method to general $D$-dimensional quadratic potentials.

32.9. **Time evolution of $R$.** (Continuation of exercise 32.6). Calculate the time evolution of $R(q,0) = a + bq + cq^2$ for a 1-dimensional harmonic oscillator using (32.12) and (32.14).

32.10. **$D$-dimensional free particle propagator.** Verify the results in sect. 32.2.2; show explicitly that (32.34), the semiclassical Van Vleck propagator in $D$ dimensions, solves the Schrödinger's equation.

32.11. **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.

32.12. **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.

32.13. **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.

32.14. **Zero length orbits.** Derive the classical trace (16.1) rigorously and either add the $t \to 0_+$ zero length contribution to the trace formula, or show that it vanishes. Send us a reprint of *Phys. Rev. Lett.* with the correct derivation.

32.15. **Free particle semiclassical Green's functions.** Calculate the semiclassical Green's functions for the systems of exercise 32.13.

## References

[32.1] A. Einstein, "On the Quantum Theorem of Sommerfeld and Epstein," p. 443, English translation of "Zum Quantensatz von Sommerfeld und Epstein," *Verh. Deutsch. Phys. Ges.* **19**, 82 (1917), in *The Collected Papers of Albert Einstein*, Volume **6**: *The Berlin Years: Writings, 1914-1917*, A. Engel, transl. and E. Schucking, (Princeton University Press, Princeton, New Jersey 1997).

[32.2] E. Madelung, *Zeitschr. fr Physik* **40**, 332 (1926).

[32.3] E. Schrödinger, *Annalen der Physik* **79**, 361, 489; **80**, 437, **81**, 109 (1926).

[32.4] J. H. van Vleck, *Quantum Principles and the Line Spectra*, Bull. Natl. Res. Council **10**, 1 (1926).

[32.5] J. H. van Vleck, *Proc. Natl. Acad. Sci.* **14**, 178 (1928).

[32.6] D. Bohm, *Phys. Rev.* **85**, 166 (1952).

[32.7] P.R. Holland, *The quantum theory of motion - An account of the de Broglie-Bohm casual interpretation of quantum mechanics* (Cambridge Univ. Press, Cambridge 1993).

[32.8] physics.nist.gov/cgi-bin/cuu

# Chapter 33

# 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 classically chaotic systems. In chapter 34 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 37.

## 33.1 Trace formula

Our next task is to evaluate the Green's function trace (30.17) in the semiclassical approximation. The trace

$$\operatorname{tr} G_{sc}(E) = \int d^D q\, G_{sc}(q,q,E) = \operatorname{tr} G_0(E) + \sum_j \int d^D q\, G_j(q,q,E)$$

receives contributions from "long" classical trajectories labeled by $j$ which start and end in $q$ after finite time, and the "zero length" trajectories whose lengths approach zero as $q' \to q$.

First, we work out the contributions coming from the finite time *returning* classical 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 $q' \to q$ limit,

$$\left.\frac{\partial S_j(q,q',E)}{\partial q_i}\right|_{q'=q} + \left.\frac{\partial S_j(q,q',E)}{\partial q'_i}\right|_{q'=q} = 0\,,$$

**Figure 33.1:** 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.

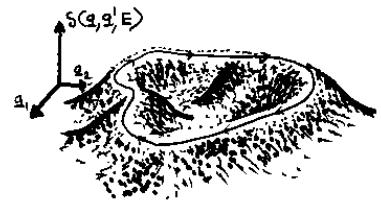



**Figure 33.2:** A romanticized sketch of $S_p(E) = S(q,q,E) = \oint p(q,E)dq$ landscape orbit. Unstable periodic orbits traverse isolated ridges and saddles of the mountainous landscape of the action $S(q_\parallel, q_\perp, E)$. Along a periodic orbit $S_p(E)$ is constant; in the transverse directions it generically changes quadratically.

meaning that the initial and final momenta (32.40) of contributing trajectories should coincide

$$p_i(q,q,E) - p'_i(q,q,E) = 0\,, \qquad q \in j\text{th periodic orbit}\,, \tag{33.1}$$

so the trace receives contributions only from those long classical trajectories which are *periodic* in the full phase space.

For a periodic orbit the natural coordinate system is the intrinsic one, with $q_\parallel$ 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

$$\mathrm{tr}\, G_j(E) = \frac{1}{i\hbar(2\pi\hbar)^{(d-1)/2}} \oint_j \frac{dq_\parallel}{\dot{q}} \int_j d^{d-1}q_\perp |\det D^j_\perp|^{1/2} e^{\frac{i}{\hbar}S_j - \frac{i\pi}{2}m_j}\,,$$

where the integration in $q_\parallel$ 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 $S_j(q_\parallel, q_\perp, E)$, and assume that the density varies smoothly and is well approximated by its value $D^j_\perp(q_\parallel, 0, E)$ on the classical trajectory, $q_\perp = 0$ . The topological index $m_j(q_\parallel, q_\perp, E)$ is an integer which does not depend on the initial point $q_\parallel$ and not change in the infinitesimal neighborhood of an isolated periodic orbit, so we set $m_j(E) = m_j(q_\parallel, q_\perp, E)$.

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

$$\mathrm{tr}\, G_j(E) = \frac{1}{i\hbar} \oint \frac{dq_\parallel}{\dot{q}} \left| \frac{\det D_{\perp j}(q_\parallel, 0, E)}{\det D'_{\perp j}(q_\parallel, 0, E)} \right|^{1/2} e^{\frac{i}{\hbar}S_j - \frac{i\pi}{2}m_j}\,,$$

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. Mercifully, this integral also removes most of the $2\pi\hbar$ prefactors in (??).

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

$$\begin{aligned}
\det D_\perp &= \left\|\frac{\partial p'_\perp}{\partial q_\perp}\right\| = \left\|\frac{\partial(q'_\perp, p'_\perp)}{\partial(q_\perp, q'_\perp)}\right\| \\
\det D'_\perp &= \left\|\frac{\partial p_\perp}{\partial q_\perp} - \frac{\partial p'_\perp}{\partial q_\perp} + \frac{\partial p_\perp}{\partial q'_\perp} - \frac{\partial p'_\perp}{\partial q'_\perp}\right\| = \left\|\frac{\partial(p_\perp - p'_\perp, q_\perp - q'_\perp),}{\partial(q_\perp, q'_\perp)}\right\| .
\end{aligned}$$

Defining the $2(D-1)$-dimensional transverse vector $x_\perp = (q_\perp, p_\perp)$ in the full phase space we can express the ratio

$$\begin{aligned}
\frac{\det D'_\perp}{\det D_\perp} &= \left\|\frac{\partial(p_\perp - p'_\perp, q_\perp - q'_\perp)}{\partial(q'_\perp, p'_\perp)}\right\| = \left\|\frac{\partial(x_\perp - x'_\perp)}{\partial x'_\perp}\right\| \\
&= \det\,(M - \mathbf{1})\ , \qquad (33.2)
\end{aligned}$$

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 classical periodic orbit action $S_j(E) = \oint p(q_\|, E)dq_\|$ is an integral around a loop defined by the periodic orbit, and does not depend on the starting point $q_\|$ along the orbit, see figure 33.2. 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 the $q_\|$ integral

$$\operatorname{tr} G_j(E) = \frac{1}{i\hbar}\sum_j \frac{1}{|\det(1-M_j)|^{1/2}} e^{r(\frac{i}{\hbar}S_j - \frac{i\pi}{2}m_j)} \oint \frac{dq_\|}{\dot{q}_\|} .$$

Here we have assumed that $M_j$ has no marginal eigenvalues. The determinant of the monodromy matrix, the action $S_p(E) = \oint p(q_\|, E)dq_\|$ and the topological index are all classical 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 $dt = dq_{\|}/\dot{q}(t)$

$$\int_0^{L_p} \frac{dq_{\|}}{\dot{q}(t)} = \int_0^{T_p} dt = T_p \,.$$

Note that the spatial integral corresponds to a *single* traversal. If you do not see why this is so, rethink the derivation of the classical trace formula (16.23) - 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*

$$\mathrm{tr}\, G_{sc}(E) = \mathrm{tr}\, G_0(E) + \frac{1}{i\hbar} \sum_p T_p \sum_{r=1}^{\infty} \frac{1}{|\det(1 - M_p^r)|^{1/2}} e^{r(\frac{i}{\hbar} S_p - \frac{i\pi}{2} m_p)} \,, \qquad (33.3)$$

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 $(q_i, p_i)$, $i = 1, 2, \ldots, D$ (no sum on $i$) change their signs as one goes once around the curve.

### 33.1.1 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 $q' \to q$ 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 (30.17) is given by the imaginary part of the Gutzwiller trace formula (33.3) multiplied with $-1/\pi$. The contribution coming from the zero length trajectories is the imaginary part of (32.48) for $q' \to q$ integrated over the configuration space

$$d_0(E) = -\frac{1}{\pi} \int d^D q \, \mathrm{Im}\, \mathrm{G}_0(\mathrm{q}, \mathrm{q}, \mathrm{E}),$$

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,

$$N_{sc}(E) = \frac{1}{h^D} \int d^D p d^D q \, \Theta(E - H(q,p)) \,. \tag{33.4}$$

where $\Theta(x)$ is the Heaviside function (30.22). $N_{sc}(E)$ is an estimate of the spectral staircase (30.21), so its derivative yields the average density of states

$$d_0(E) = \frac{d}{dE} N_{sc}(E) = \frac{1}{h^D} \int d^D p d^D q \, \delta(E - H(q,p)) \,, \tag{33.5}$$

precisely the semiclassical result (33.6). For Hamiltonians of type $p^2/2m + V(q)$, the energy shell volume in (33.5) 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 33.3]

$$d_0(E) = \frac{2m}{\hbar^D 2^d \pi^{D2} \Gamma(D/2)} \int_{V(q)<E} d^D q \, [2m(E - V(q))]^{D/2-1} \,, \tag{33.6}$$

and

$$N_{sc}(E) = \frac{1}{h^D} \frac{\pi^{D/2}}{\Gamma(1 + D/2)} \int_{V(q)<E} d^D q \, [2m(E - V(q))]^{D/2} \,. \tag{33.7}$$

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 33.4]

$$d_0(E) = \frac{T(E)}{2\pi\hbar} \,, \tag{33.8}$$

where $T(E)$ is the period of the periodic orbit of fixed energy $E$. In two spatial dimensions the average density of states is

$$d_0(E) = \frac{m\mathcal{A}(E)}{2\pi\hbar^2} \,, \tag{33.9}$$

where $\mathcal{A}(E)$ is the classically allowed area of configuration space for which $V(q) < E$.

[exercise 33.5]

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

$$d_{osc}(E) = \frac{1}{\pi\hbar} \sum_p T_p \sum_{r=1}^{\infty} \frac{\cos(rS_p(E)/\hbar - rm_p\pi/2)}{|\det(1 - M_p^r)|^{1/2}} \tag{33.10}$$

follows from the trace formula (33.3).

### 33.1.2 Regularization of the trace

The real part of the $q' \to q$ zero length Green's function (32.48) is ultraviolet divergent in dimensions $d > 1$, and so is its formal trace (30.17). The short distance behavior of the real part of the Green's function can be extracted from the real part of (32.48) 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 $\gamma = 0.577...$ is the Euler constant. The real part of the Green's function for short distance is dominated by the singular part

$$G_{sing}(|q-q'|,E) = \begin{cases} -\frac{m}{2\hbar^2\pi^{\frac{d}{2}}}\Gamma((d-2)/2)\frac{1}{|q-q'|^{d-2}} & \text{for } d \neq 2 \\ \frac{m}{2\pi\hbar^2}(\ln(2m(E-V)|q-q'|/2\hbar)+\gamma) & \text{for } d = 2 \end{cases}.$$

The *regularized* Green's function

$$G_{reg}(q,q',E) = G(q,q',E) - G_{sing}(|q-q'|,E)$$

is obtained by subtracting the $q' \to q$ ultraviolet divergence. For the regularized Green's function the Gutzwiller trace formula is

$$\operatorname{tr} G_{reg}(E) = -i\pi d_0(E) + \frac{1}{i\hbar}\sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r(\frac{i}{\hbar}S_p(E)-\frac{i\pi}{2}m_p(E))}}{|\det(1-M_p^r)|^{1/2}}. \tag{33.11}$$

Now you stand where Gutzwiller stood in 1990. You hold the trace formula in your hands. You have no clue how good is the $\hbar \to 0$ approximation, how to take care of the sum over an infinity of periodic orbits, and whether the formula converges at all.

## 33.2 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 (30.18) 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.

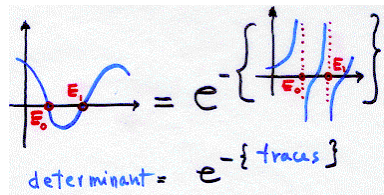



**Figure 33.3:** A sketch of how spectral determinants convert poles into zeros: The trace shows $1/(E - E_n)$ type singularities at the eigenenergies while the spectral determinant goes smoothly through zeroes.

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 33.1.2. 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

$$-\frac{d}{dE} \ln \det(\hat{H} - E) = \sum_n \frac{1}{E - E_n} = \operatorname{tr} G(E).$$

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

$$-\frac{d}{dE} \ln \det(\hat{H} - E)_{sc} = \operatorname{tr} G_{reg}(E).$$

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 33.3. By integrating and exponentiating we obtain

$$\det(\hat{H} - E)_{sc} = \exp\left(-\int^E dE' \operatorname{tr} G_{reg}(E')\right)$$

Now we can use (33.11) and integrate the terms coming from periodic orbits, using the relation (32.17) between the action and the period of a periodic orbit, $dS_p(E) = T_p(E)dE$, and the relation (30.21) between the density of states and the spectral staircase, $dN_{sc}(E) = d_0(E)dE$. We obtain the *semiclassical zeta function*

$$\det(\hat{H} - E)_{sc} = e^{i\pi N_{sc}(E)} \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{ir(S_p/\hbar - m_p\pi/2)}}{|\det(1 - M_p^r)|^{1/2}}\right). \tag{33.12}$$

[chapter 18]

We already know from the study of classical evolution operator spectra of chapter 17 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 $p$.

## 33.3 One-dof 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-dof form of the oscillating density (33.10) and of the average density (33.8)

$$d(E) = \frac{T_p(E)}{2\pi\hbar} + \sum_r \frac{T_p(E)}{\pi\hbar} \cos(rS_p(E)/\hbar - rm_p(E)\pi/2). \tag{33.13}$$

The classical 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 (33.13) 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

$$d(E) = \frac{T_p(E)}{2\pi\hbar} \sum_n \delta(S_p(E)/2\pi\hbar - m_p(E)/4 - n). \tag{33.14}$$

This expression can be simplified by using the relation (32.17) between $T_p$ and $S_p$, and the identity (14.7) $\delta(x-x^*) = |f'(x)|\delta(f(x))$, 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 (33.14)

$$S_p(E_n)/2\pi\hbar = n - m_p(E)/4\,,$$

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

$$\oint p(q, E_n)dq = h\left(n - \frac{m_p}{4}\right). \tag{33.15}$$

In this way the trace formula recovers the well known 1-dof 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

$$N_{sc}(E) = n - 1/2 = S_p(E)/2\pi\hbar - m_p(E)/4 - 1/2. \tag{33.16}$$

The 1-dof spectral determinant follows from (33.12) by dropping the monodromy matrix part and using (33.16)

$$\det(\hat{H} - E)_{sc} = \exp\left(-\frac{i}{2\hbar}S_p + \frac{i\pi}{2}m_p\right)\exp\left(-\sum_r \frac{1}{r}e^{\frac{i}{\hbar}rS_p - \frac{i\pi}{2}rm_p}\right). \tag{33.17}$$

Summation yields a logarithm by $\sum_r t^r/r = -\ln(1-t)$ and we get

$$\begin{aligned}\det(\hat{H} - E)_{sc} &= e^{-\frac{i}{2\hbar}S_p + \frac{im_p}{4} + \frac{i\pi}{2}}(1 - e^{\frac{i}{\hbar}S_p - i\frac{m_p}{2}}) \\ &= 2\sin\left(S_p(E)/\hbar - m_p(E)/4\right).\end{aligned}$$

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 (33.15).

## 33.4 Two-dof systems

For flows in two configuration dimensions the monodromy matrix $M_p$ has two eigenvalues $\Lambda_p$ and $1/\Lambda_p$, as explained in sect. 7.2. 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-dof expression for the average density of states (33.9) the spectral determinant becomes

$$\begin{aligned}\det(\hat{H}-E)_{sc} &= e^{i\frac{m\mathcal{A}E}{2\hbar^2}}\exp\left(-\sum_p\sum_{r=1}^{\infty}\sum_{k=0}^{\infty}\frac{e^{ir(S_p/\hbar-m_p\pi/2)}}{r|\Lambda_p^r|^{1/2}\Lambda_p^{kr}}\right)\\ &= e^{i\frac{m\mathcal{A}E}{2\hbar^2}}\prod_p\prod_{k=0}^{\infty}\left(1-\frac{e^{\frac{i}{\hbar}S_p-\frac{i\pi}{2}m_p}}{|\Lambda_p|^{1/2}\Lambda_p^k}\right). \qquad (33.18)\end{aligned}$$

## Résumé

Spectral determinants and dynamical zeta functions arise both in classical 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 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 *classical* in the sense that all quantities used in periodic orbit calculations - actions, stabilities, geometrical phases - are classical quantities. The problem is then to understand and control the convergence of classical 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

$$\operatorname{tr} G_{sc}(E) = \int dq\, G_{sc}(q,q,E)$$

is given by a sum over the periodic orbits (33.11). 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 classical dynamics trace formulas hide under a variety of appellations such as the $f-\alpha$ 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 classical 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-dof semiclassical spectral determinant (33.18)

$$\det(\hat{H}-E)_{sc} = e^{i\pi N_{sc}(E)} \prod_p \prod_{k=0}^{\infty} \left(1 - \frac{e^{iS_p/\hbar - i\pi m_p/2}}{|\Lambda_p|^{1/2}\Lambda_p^k}\right)$$

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 33.1** Gutzwiller quantization of classically chaotic systems. The derivation given here and in sects. 32.3 and 33.1 follows closely the excellent exposition [2] by Martin Gutzwiller, the inventor of the trace formula. The derivation presented here is self contained, but refs. [3, 1] might also be of help to the student.

**Remark 33.2** Zeta functions. For "zeta function" nomenclature, see remark 17.4 on page 296.

## Exercises

33.1. **Monodromy matrix from second variations of the action.** Show that

$$D_{\perp j}/D'_{\perp j} = (\mathbf{1} - M) \tag{33.19}$$

33.2. **Jacobi gymnastics.** Prove that the ratio of determinants in (S.48) can be expressed as

$$\frac{\det D'_{\perp j}(q_{\|}, 0, E)}{\det D_{\perp j}(q_{\|}, 0, E)} = \det \begin{pmatrix} I - M_{qq} & -M_{qp} \\ -M_{pq} & I - M_{pp} \end{pmatrix} = \det(1 - M_j)\,, \tag{33.20}$$

where $M_j$ is the monodromy matrix of the periodic orbit.

33.3. **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 $\Gamma(1 + d/2) = \Gamma(d/2)d/2$.

33.4. **Average density of states in 1 dimension.** Show that in one dimension the average density of states is given by (33.8)

$$\bar{d}(E) = \frac{T(E)}{2\pi\hbar},$$

where $T(E)$ is the time period of the 1-dimensional motion and show that

$$\bar{N}(E) = \frac{S(E)}{2\pi\hbar}\,, \tag{33.21}$$

where $S(E) = \oint p(q, E)\, dq$ is the action of the orbit.

33.5. **Average density of states in 2 dimensions.** Show that in 2 dimensions the average density of states is given by (33.9)

$$\bar{d}(E) = \frac{m\mathcal{A}(E)}{2\pi\hbar^2}\,,$$

where $\mathcal{A}(E)$ is the classically allowed area of configuration space for which $U(q) < E$.

## References


[33.1] R.G. Littlejohn, *J. Stat. Phys.* **68**, 7 (1992).

[33.2] L.D. Landau and E.M. Lifshitz, *Mechanics* (Pergamon, London, 1959).

[33.3] R.G. Littlejohn, "Semiclassical structure of trace formulas," in G. Casati and B. Chirikov, eds., *Quantum Chaos*, (Cambridge University Press, Cambridge 1994).

[33.4] M.C. Gutzwiller, *J. Math. Phys.* **8**, 1979 (1967); **10**, 1004 (1969); **11**, 1791 (1970); **12**, 343 (1971).

[33.5] M.C. Gutzwiller, *J. Math. Phys.* **12**, 343 (1971)

[33.6] M.C. Gutzwiller, *J. Phys. Chem.* **92**, 3154 (1984).

[33.7] A. Voros, *J. Phys.* **A 21**, 685 (1988).

[33.8] A. Voros, *Aspects of semiclassical theory in the presence of classical chaos*, *Prog. Theor. Phys. Suppl.* **116**, 17 (1994).

[33.9] P. Cvitanović and P.E. Rosenqvist, in G.F. Dell'Antonio, S. Fantoni and V.R. Manfredi, eds., *From Classical to Quantum Chaos, Soc. Italiana di Fisica Conf. Proceed.* **41**, pp. 57-64 (Ed. Compositori, Bologna 1993).

[33.10] A. Wirzba, "Validity of the semiclassical periodic orbit approximation in the 2-and 3-disk problems," *CHAOS* **2**, 77 (1992).

[33.11] P. Cvitanović, G. Vattay and A. Wirzba, 'Quantum fluids and classical determinants," in H. Friedrich and B. Eckhardt., eds., *Classical, Semiclassical and Quantum Dynamics in Atoms – in Memory of Dieter Wintgen, Lecture Notes in Physics* **485** (Springer, New York 1997), `chao-dyn/9608012`.

[33.12] E.B. Bogomolny, CHAOS **2**, 5 (1992).

[33.13] E.B. Bogomolny, *Nonlinearity* **5**, 805 (1992).

[33.14] M. Kline, *Mathematical Thought from Ancient to Modern Times* (Oxford Univ. Press, Oxford 1972); on Monge and theory of characteristics - chapter 22.7.

[33.15] E.T. Bell, *Men of Mathematics* (Penguin, London 1937).

[33.16] R.P. Feynman, *Statistical Physics* (Addison Wesley, New York 1990).

[33.17] H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, 1980); chapter 9.

[33.18] G. Tanner and D. Wintgen, *CHAOS* **2**, 53 (1992).

[33.19] P. Cvitanović and F. Christiansen, *CHAOS* **2**, 61 (1992).

[33.20] M.V. Berry and J.P. Keating, *J. Phys.* **A 23**, 4839 (1990).

[33.21] H.H. Rugh, "Generalized Fredholm determinants and Selberg zeta functions for Axiom A dynamical systems," *Ergodic Theory Dynamical Systems* **16**, 805 (1996).

[33.22] B. Eckhard and G. Russberg, Phys. Rev. **E 47**, 1578 (1993).

[33.23] D. Ruelle, *Statistical Mechanics, Thermodynamical Formalism* (Addison-Wesley, Reading MA, 1987).

[33.24] P. Szépfalusy, T. Tél, A. Csordás and Z. Kovács, Phys. Rev. **A 36**, 3525 (1987).

[33.25] H.H. Rugh, Nonlinearity **5**, 1237 (1992) and H.H. Rugh, *Ph.D. Thesis* (Niels Bohr Institute, 1993).

[33.26] P. Cvitanović, P.E. Rosenqvist, H.H. Rugh and G. Vattay, *Scattering Theory - special issue*, CHAOS (1993).

[33.27] E.J. Heller, S. Tomsovic and A. Sepúlveda CHAOS **2**, *Periodic Orbit Theory - special issue*, 105, (1992).

[33.28] V.I. Arnold, *Geometrical Methods in the Theory of Ordinary Differential Equations*, (Springer, New York 1983).

[33.29] R. Dashen, B. Hasslacher and A. Neveu , "Nonperturbative methods and extended hadron models in field theory. 1. Semiclassical functional methods," *Phys. Rev.* **D10**, 4114 (1974).

[33.30] V.I. Arnold, *Geometrical Methods in the Theory of Ordinary Differential Equations* (Springer, New York 1983).

# Chapter 34

# 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 function, which we learned to analyze in earlier chapters.

## 34.1 Density of states

For a scattering problem the density of states (30.18) appear ill defined since formulas such as (33.6) 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 (30.5)

$$\left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial q^2} + V(q)\right)\phi_{\vec{k}}(q) = E\phi_{\vec{k}}(q) \tag{34.1}$$

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 (34.1) can be written as superposition of ingoing and outgoing solutions of the free particle Schrödinger equation for fixed angular momentum:

$$\phi(q) = A\phi^{(-)}(q) + B\phi^{(+)}(q)\,, \qquad (+\text{ boundary conditions})\,,$$

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 34.1), since the incoming ones, by definition, have yet to encounter the scattering region. This defines the quantum mechanical scattering matrix, or the *S matrix*

$$\phi_m(r) = \phi_m^{(-)}(r) + S_{mm'}\phi_{m'}^{(+)}(r)\,. \tag{34.2}$$

All scattering effects are incorporated in the deviation of $\mathbf{S}$ from the unit matrix, the transition matrix $\mathbf{T}$

$$\mathbf{S} = \mathbf{1} - i\mathbf{T}\,. \tag{34.3}$$

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 $\beta$, $\gamma$, the $S$ matrix is a collection $S_{\beta\gamma}$ of transition amplitudes $\beta \to \gamma$ normalized such that $|S_{\beta\gamma}|^2$ is the probability of the $\beta \to \gamma$ transition. The total probability that the ingoing state $\beta$ ends up in some outgoing state must add up to unity

$$\sum_\gamma |S_{\beta\gamma}|^2 = 1\,, \tag{34.4}$$

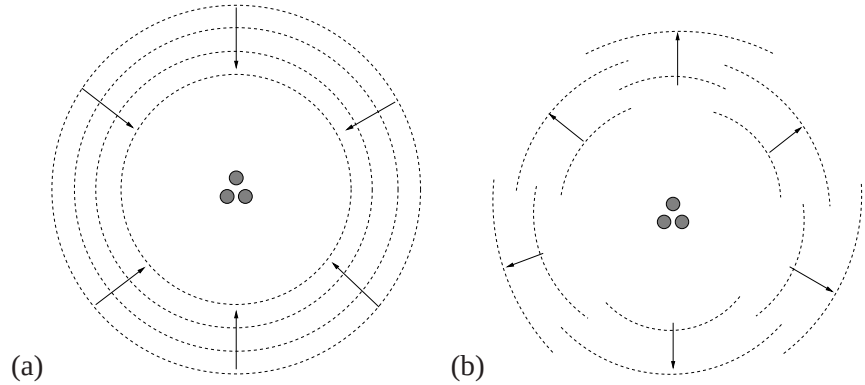



**Figure 34.1:** (a) *Incoming* spherical waves running into an obstacle. (b) Superposition of *outgoing* spherical waves scattered from an obstacle.

so the $S$ matrix is unitary: $\mathbf{S}^\dagger\mathbf{S} = \mathbf{S}\mathbf{S}^\dagger = \mathbf{1}$.

We have already encountered a solution to the 2-dimensional problem; free particle propagation Green's function (32.48) 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 (34.1) 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 (30.5) 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 (34.2) with matrix elements given by

$$S_m(k) = e^{2i\delta_m(k)}. \tag{34.5}$$

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 $R \to \infty$. 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

$$\begin{aligned}\phi_m(r) &\approx e^{im\theta}\left(S_m(k)H_m^{(+)}(kr) + H_m^{(-)}(kr)\right)\\ &\approx \cdots\cos(kr + \delta_m(k) - \chi_m)\,,\end{aligned} \tag{34.6}$$

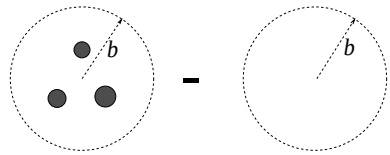



**Figure 34.2:** The "difference" of two bounded reference systems, one with and one without the scattering system.

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.

The state (34.6) must satisfy the external boundary condition that it vanish at $r = R$. This implies the quantization condition

$$k_n R + \delta_m(k_n) - \chi_m = \pi\,(n + 12)\ .$$

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. is

$$k_{n+1}R + \delta_m(k_n) + (k_{n+1} - k_n)\delta'_m(k_n) - \chi_m \approx \pi + \pi(n + 12)\,.$$

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 sbuspace

$$d_m(k) \approx \frac{1}{\pi}\,(R + \delta'_m(k))\,.$$

The $R$ term is essentially the $1-d$ Weyl term (33.8), 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 34.2. We also sum over all $m$ values so that

$$\begin{aligned} d(k) - d_0(k) = \frac{1}{\pi}\sum_m \delta'_m(k) &= \frac{1}{2\pi i}\sum_m \frac{d}{dk}\log S_m \\ &= \frac{1}{2\pi i}\mathrm{Tr}\left(S^\dagger \frac{dS}{dk}\right). \end{aligned} \tag{34.7}$$

The first line follows from the definition of the phase shifts (34.5) while the second line follows from the unitarity of $S$ so that $S^{-1} = S^\dagger$. We can now take the limit $R \to \infty$ 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,

$$d(k) - d_0(k) = -\frac{1}{2\pi k}\mathrm{Im}\left(\mathrm{tr}\left(G(k) - G_0(k)\right)\right). \tag{34.8}$$

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 (34.7) applies even when there is no circular symmetry. The proof is more difficult since one cannot appeal to the phase shifts $\delta_m$ but must work directly with a non-diagonal $S$ matrix.

## 34.2 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 $\beta$, and in the distant future they will likewise be free, in a state labeled $\gamma$. In the Heisenberg picture the $S$-matrix is defined as $\mathbf{S} = \Omega_-\Omega_+^\dagger$ in terms of the Møller operators

$$\Omega_\pm = \lim_{t\to\pm\infty} e^{iHt/\hbar}e^{-iH_0t/\hbar}\,, \tag{34.9}$$

where $H$ is the full Hamiltonian, whereas $H_0$ is the free Hamiltonian. In the interaction picture the $S$-matrix is given by

$$\begin{aligned}\mathbf{S} &= \Omega_+^\dagger\Omega_- = \lim_{t\to\infty} e^{iH_0t/\hbar}e^{-2iHt/\hbar}e^{iH_0t/\hbar}\\ &= T\exp\left(-i\int_{-\infty}^{+\infty} dtH'(t)\right)\,,\end{aligned} \tag{34.10}$$

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

$$\begin{aligned}S &= \int_0^\infty dE\, S(E)\delta(H_0 - E)\\ S(E) &= Q_+(E)Q_-^{-1}(E), \qquad Q_\pm(E) = \mathbf{1} + (H_0 - E \pm i\epsilon)^{-1}V\,,\end{aligned} \tag{34.11}$$

such that

$$\mathrm{Tr}\left[S^{\dagger}(E)\frac{d}{dE}S(E)\right] = \mathrm{Tr}\left[\frac{1}{H_0 - E - i\epsilon} - \frac{1}{H - E - i\epsilon} - (\epsilon \leftrightarrow -\epsilon)\right] . \quad (34.12)$$

The manipulations leading to (34.12) are justified if the operators $Q_{\pm}(E)$ can be linked to trace-class operators. [appendix J]

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.

## 34.3 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 32 we linked the spectral density (see (30.18)) of a bounded system

$$d(E) \equiv \sum_n \delta(E_n - E) \quad (34.13)$$

via the identity

$$\begin{aligned} \delta(E_n - E) &= -\lim_{\epsilon\to 0}\frac{1}{\pi}\,\mathrm{Im}\frac{1}{E - E_n + i\epsilon} \\ &= -\lim_{\epsilon\to 0}\frac{1}{\pi}\,\mathrm{Im}\langle E_n|\frac{1}{E - H + i\epsilon}|E_n\rangle \\ &= \frac{1}{2\pi i}\lim_{\epsilon\to 0}\left\langle E_n\left|\frac{1}{E - H - i\epsilon} - \frac{1}{E - H + i\epsilon}\right|E_n\right\rangle \end{aligned} \quad (34.14)$$

to the trace of the Green's function (33.1.1). Furthermore, in the semiclassical approximation, the trace of the Green's function is given by the Gutzwiller trace formula (33.11) 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 (34.12) provides the clue. Note that the right hand side of (34.12) has nearly the structure of (34.14) when the latter is inserted into (34.13). 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 conatiner 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:

$$\phi(\vec{r})|_{r=R} = 0 \ . \tag{34.15}$$

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 $R \to \infty$ 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 34.2). Then all the divergences linked to the increasing radius $R$ in the limit $R \to \infty$ drop out of the difference. Furthermore, in the limit $R \to \infty$ the energy-eigenfunctions of the box are only normalizable as a delta distribution, similarly to a plane wave. So we seem to recover a continous 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 $\eta$ in the energy $E$:

$$d(E+i\eta;R) \equiv \frac{1}{i\,2\pi}\sum_n \left\{ \frac{1}{E-E_n(R)-i\eta} - \frac{1}{E-E_n(R)+i\eta} \right\} \ . \tag{34.16}$$

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 $\eta$ 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 $\eta \to 0$ can be carried out, respecting the order of the limiting procedures. First, the limit $R \to \infty$ has to be performed for a *finite* value of $\eta$, only then the limit $\eta \to 0$ 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:

$$\begin{aligned}\lim_{\eta\to+0}\lim_{R\to\infty}\Big(d(E+i\eta;R)-d^{(0)}(E+i\eta;R)\Big) &= \frac{1}{2\pi i}\mathrm{Tr}\left[S^{\dagger}(E)\frac{d}{dE}S(E)\right]\\ = \frac{1}{2\pi i}\mathrm{Tr}\frac{d}{dE}\ln S(E) &= \frac{1}{2\pi i}\frac{d}{dE}\ln\det S(E)\,. \qquad (34.17)\end{aligned}$$

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 34.2). 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 (33.10) and an oscillating sum over periodic orbits. As in (33.2), the trace formula (34.17) can be integrated to give a relation between the smoothed staircase functions and the determinant of the $S$-matrix:

$$\lim_{\eta\to+0}\lim_{R\to\infty}\Big(N(E+i\eta;R)-N^{(0)}(E+i\eta;R)\Big) = \frac{1}{2\pi i}\ln\det S(E)\,. \qquad (34.18)$$

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, $\eta'$ must be greater than zero. Otherwise, the exact left hand sides (34.18) and (34.17) 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 (34.18) and (34.17) 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 $R\to\infty$. The periodic orbits that encounter both the scattering region and the external wall would still survive the first limit $R\to\infty$, 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 (34.17) and (34.18) are evaluated at a finite value of $R$.

[exercise 34.1]

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.

## 34.4 Wigner time delay

The term $\frac{d}{dE}\ln\det S$ in the density formula (34.17) is dimensionally time. This suggests another, physically important interpretation of such formulas for scattering systems, the Wigner delay, defined as

$$\begin{aligned} d(k) &= \frac{d}{dk}\mathrm{Argdet}\,(\mathbf{S}(k)) \\ &= -i\frac{d}{dk}\log\det\,(\mathbf{S}(k) \\ &= -i\,\mathrm{tr}\left(\mathbf{S}^{\dagger}(k)\frac{d\mathbf{S}}{dk}(k)\right) \end{aligned} \tag{34.19}$$

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:

$$\phi = \vec{k}\cdot\vec{x} - \omega\,t \;+\; \Theta\,.$$

Here the term in the parenthesis refers to the phase shift that will occur if scattering is present. 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 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

$$\Delta t_{ij} = \mathrm{Re}\left(-\mathrm{i}\,\mathrm{S}_{\mathrm{ij}}^{-1}\frac{\partial \mathrm{S}_{\mathrm{ij}}}{\partial\omega}\right) = \mathrm{Re}\left(\frac{\partial\Theta_{\mathrm{ij}}}{\partial\omega}\right)$$

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

$$\begin{aligned}\langle\Delta t_i\rangle &= \sum_j |S_{ij}|^2\Delta t_{ij} = \mathrm{Re}\,(-\mathrm{i}\sum_{\mathrm{j}}\mathrm{S}_{\mathrm{ij}}^{*}\frac{\partial \mathrm{S}_{\mathrm{ij}}}{\partial\omega}) = \mathrm{Re}\,(-\mathrm{i}\,\mathbf{S}^{\dagger}\cdot\frac{\partial\mathbf{S}}{\partial\omega})_{\mathrm{ii}} \\ &= -i\left(\mathbf{S}^{\dagger}\cdot\frac{\partial\mathbf{S}}{\partial\omega}\right)_{ii}\,,\end{aligned}$$

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 (34.19).

## Commentary

**Remark 34.1** Krein-Friedel-Lloyd formula. The third volume of Thirring [1], 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 the book of K. Huang, chapter 10 (on the "second virial coefficient"), or Beth and Uhlenbeck [5], or Friedel [7]. These results for the correction to the density of states are particular cases of the Krein formula [3]. The Krein-Friedel-Lloyd formula (34.17) was derived in refs. [3, 7, 8, 9], see also refs. [11, 14, 15, 17, 18]. The original papers are by Krein and Birman [3, 4] but beware, they are mathematicans.

Also, have a look at pages 15-18 of Wirzba's talk on the Casimir effect [16]. Page 16 discusses the Beth-Uhlenbeck formula [5], the predecessor of the more general Krein formula for spherical cases.

**Remark 34.2** 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 34.3** 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 [26], they are the sum over all ouput channels (both in reflection and transmission) of $\Delta t_{ab} = \mathrm{Re}\,[(-\mathrm{i}/\mathrm{S}_{\mathrm{ab}})(\partial \mathrm{S}_{\mathrm{ab}}/\partial\omega)]$ 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.

## Exercises

34.1. **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 $\eta \to 0$, container size $b \to \infty$?

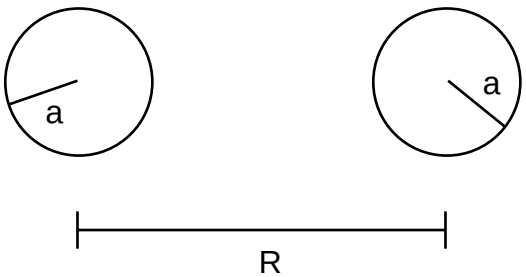
34.2. **The one-disk scattering wave function.** Derive the one-disk scattering wave function.

(Andreas Wirzba)

34.3. **Quantum two-disk scattering.** Compute the quasi-classical 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

$$M_{m,n} = \delta_{m,n} + \frac{(-1)^n}{2} \frac{J_m(ka)}{H_n^{(1)}(ka)} \left( H_{m-n}^{(1)}(kR) + (-1)^n H_{m+n}^{(1)}(\right.$$

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: notice the structure $M = I + A$ to approximate the determinant; or read *Chaos* **2**, 79 (1992))

34.4. **Pinball topological index.** Upgrade your pinball simulator so that it computes the topological index for each orbit it finds.

## References

[34.1] W. Thirring, *Quantum mechanics of atoms and molecules, A course in mathematical physics* Vol. **3** (Springer, New York, 1979). (Springer, Wien 1979).

[34.2] A. Messiah, *Quantum Mechanics, Vol.* **I** (North-Holland, Amsterdam, 1961).

[34.3] M.G. Krein, *On the Trace Formula in Perturbation Theory*, Mat. Sborn. (N.S.) **33**, 597 (1953) ; *Perturbation Determinants and Formula for Traces of Unitary and Self-adjoint Operators*, Sov. Math.-Dokl. **3**, 707 (1962).

[34.4] M.Sh. Birman and M.G. Krein, *On the Theory of Wave Operators and Scattering Operators*, Sov. Math.-Dokl. **3**, 740 (1962); M.Sh. Birman and D.R. Yafaev, *St. Petersburg Math. J.* **4**, 833 (1993).

[34.5] E. Beth and G.E. Uhlenbeck, *Physica* **4**, 915 (1937).

[34.6] K. Huang, *Statistical Mechanics* (John Wiley & Sons, New York (1987).

[34.7] J. Friedel, *Phil. Mag.* **43**, 153 (1952); *Nuovo Cim. Ser. 10 Suppl.* **7**, 287 (1958).

[34.8] P. Lloyd, Wave propagation through an assembly of spheres. II. The density of single-particle eigenstates, *Proc. Phys. Soc.* **90**, 207 (1967).

[34.9] P. Lloyd and P.V. Smith, Multiple-scattering theory in condensed materials, *Adv. Phys.* **21**, 69 (1972), and references therein.

[34.10] R. Balian and C. Bloch, *Ann. Phys. (N.Y.)* **63**, 592 (1971)

[34.11] R. Balian and C. Bloch, *Solution of the Schrödinger Equation in Terms of Classical Paths Ann. Phys. (NY)* **85**, 514 (1974).

[34.12] R. Balian and C. Bloch, *Distribution of eigenfrequencies for the wave equation in a finite domain: III. Eigenfrequency density oscillations*, *Ann. Phys. (N.Y.)* **69**,76 (1972).

[34.13] J.S. Faulkner, "Scattering theory and cluster calculations," *J. Phys.* **C 10**, 4661 (1977).

[34.14] P. Gaspard and S.A. Rice, Semiclassical quantization of the scattering from a classically chaotic repeller, *J. Chem. Phys.* **90**, 2242 (1989).

[34.15] P. Scherer, *Quantenzustände eines klassisch chaotischen Billards*, Ph.D. thesis, Univ. Köln (Berichte des Forschungszentrums Jülich 2554, ISSN 0366-0885, Jülich, Nov. 1991).

[34.16] A. Wirzba, "A force from nothing into nothing: Casimir interactions" ChaosBook.org/projects/Wirzba/openfull.ps.gz (overheads, 2003).

[34.17] P. Gaspard, Scattering Resonances: Classical and Quantum Dynamics, in: *Proceedings of the Int. School of Physics "Enrico Fermi"*, Course CXIX, Varena, 23 July - 2 August 1991, eds G. Casati, I. Guarneri and U. Smilansky (North-Holland, Amsterdam, 1993).

[34.18] A. Norcliffe and I. C. Percival, J. Phys. B **1**, 774 (1968); L. Schulman, Phys. Rev. **176**, 1558 (1968).

[34.19] W. Franz, *Theorie der Beugung Elektromagnetischer Wellen* (Springer, Berlin 1957); "Über die Greenschen Funktionen des Zylinders und der Kugel," *Z. Naturforschung* **9a**, 705 (1954).

[34.20] G.N. Watson, *Proc. Roy. Soc. London Ser.* **A 95**, 83 (1918).

[34.21] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, (Dover, New York, 1964).

[34.22] W. Franz and R. Galle, "Semiasymptotische Reihen für die Beugung einer ebenen Welle am Zylinder," *Z. Naturforschung* **10a**, 374 (1955).

[34.23] A. Wirzba, "Validity of the semiclassical periodic orbit approximation in the 2-and 3-disk problems," *CHAOS* **2**, 77 (1992).

[34.24] M.V. Berry, "Quantizing a Classically Ergodic System: Sinai's Billiard and the KKR Method," *Ann. Phys. (N.Y.)* **131**, 163 (1981).

[34.25] E.P. Wigner, *Phys. Rev.* **98**, 145 (1955).

[34.26] F.T. Smith, *Phys. Rev.* **118**, 349 (1960).

[34.27] A. Wirzba, *Quantum Mechanics and Semiclassics of Hyperbolic n-Disk Scattering*, *Physics Reports* **309**, 1-116 ( 1999); chao-dyn/9712015.

[34.28] V. A. Gopar, P. A. Mello, and M. Buttiker, Phys. Rev. Lett. 77, 3005 (1996).

[34.29] P. W. Brouwer, K. M. Frahm, and C. W. J. Beenakker, Phys. Rev. Lett. 78, 4737 (1997).

[34.30] Following the thesis of Eisenbud, the local delay time Dtab is defined in ref. [26] as the appearance of the peak in the outgoing signal in channel b after the injection of a wave packet in channel a. Our definition of the local delay time tab in Eq. (1) coincides with the definition of Dtab in the limit of narrow bandwidth pulses, as shown in Eq. (3).

[34.31] E. Doron and U. Smilansky, Phys. Rev. Lett. 68, 1255 (1992).

[34.32] G. Iannaccone, Phys. Rev. B 51, 4727 (1995).

[34.33] V. Gasparian, T. Christen, and M. Büttiker, Phys. Rev. A 54, 4022 (1996).

[34.34] For a complete and insightful review see Y. V. Fyodorv and H.-J. Sommers, J. Math. Phys. 38, 1918 (1997).

[34.35] R. Landauer and Th. Martin, Rev. Mod. Phys. 66, 217 (1994). j

[34.36] E. H. Hauge and J. A. Støveng, Rev. Mod. Phys. 61, 917 (1989).

# Chapter 35

# 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 36.

- For bound-state problem the semiclassical approximation does not respect quantum-mechanical unitarity, and the semi-classical 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 classical escape processes discussed in chapter1.
- 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.

## 35.1 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 par-

ticle moves freely between the static scatterers, the time independent Schrödinger equation outside the scattering regions is the Helmholtz equation:

$$\left(\vec{\nabla}_r^2 + \vec{k}^2\right)\psi(\vec{r}) = 0\,, \qquad \vec{r} \text{ outside the scattering regions.} \tag{35.1}$$

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:

$$\psi(\vec{r}) = 0\,, \qquad \vec{r} \text{ on the billiard perimeter} \tag{35.2}$$

As usual for scattering problems, we expand the wave function $\psi(\vec{r})$ in the (2-dimensional) angular momentum eigenfunctions basis

$$\psi(\vec{r}) = \sum_{m=-\infty}^{\infty} \psi_m^k(\vec{r})e^{-im\Phi_k}\,, \tag{35.3}$$

where $k$ and $\Phi_k$ are the length and angle of the wave vector, respectively. A plane wave in two dimensions expaned in the angular momentum basis is

$$e^{i\vec{k}\cdot\vec{r}} = e^{ikr\cos(\Phi_r-\Phi_k)} = \sum_{m=-\infty}^{\infty} J_m(kr)e^{im(\Phi_r-\Phi_k)}\,, \tag{35.4}$$

where $r$ and $\Phi_r$ 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

$$J_m(z) = \frac{1}{2}\left(H_m^{(1)}(z) + H_m^{(2)}(z)\right) \tag{35.5}$$

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:

$$\begin{aligned} H_m^{(2)}(z) &\sim \sqrt{\frac{2}{\pi z}}e^{-i(z-\frac{\pi}{2}m-\frac{\pi}{4})} \quad \text{incoming,} \\ H_m^{(1)}(z) &\sim \sqrt{\frac{2}{\pi z}}e^{+i(z-\frac{\pi}{2}m-\frac{\pi}{4})} \quad \text{outgoing.} \end{aligned} \tag{35.6}$$

Thus for $r \to \infty$ 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:

$$J_m(kr)e^{im\Phi_r} \sim \frac{1}{\sqrt{2\pi kr}}\left[e^{-i(kr-\frac{\pi}{2}m-\frac{\pi}{4})} + e^{i(kr-\frac{\pi}{2}m-\frac{\pi}{4})}\right]e^{im\Phi_r}\,. \tag{35.7}$$

In terms of the asymptotic (angular momentum) components $\psi_m^k$ of the wave function $\psi(\vec{r})$, the scattering matrix (34.3) is defined as

$$\psi_m^k \sim \frac{1}{\sqrt{2\pi kr}} \sum_{m'=-\infty}^{\infty} \left[\delta_{mm'} e^{-i(kr-\frac{\pi}{2}m'-\frac{\pi}{4})} + S_{mm'} e^{i(kr-\frac{\pi}{2}m'-\frac{\pi}{4})}\right] e^{im'\Phi_r} \ . \tag{35.8}$$

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 $\psi(\vec{r})$.

### 35.1.1 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 (35.5) the $m$th angular component of the time-independent scattering wave function is a superposition of incoming and outgoing 2-dimensional spherical waves

[exercise 34.2]

$$\begin{aligned}\psi_m^k &= \frac{1}{2}\left(H_m^{(2)}(kr) + S_{mm} H_m^{(1)}(kr)\right) e^{im\Phi_r} \\ &= \left(J_m(kr) - \frac{i}{2} T_{mm} H_m^{(1)}(kr)\right) e^{im\Phi_r} \ .\end{aligned}$$

The vanishing (35.2) 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:

$$S^s_{mm'}(k) = \left(1 - \frac{2J_m(ka_s)}{H_m^{(1)}(ka_s)}\right)\delta_{mm'} = -\frac{H_m^{(2)}(ka_s)}{H_m^{(1)}(ka_s)}\delta_{mm'} \ , \tag{35.9}$$

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. 35.3.

### 35.1.2 Multi-scattering matrix

Consider next a scattering region consisting of $N$ non-overlapping disks labeled $s \in \{1, 2, \cdots, N\}$, following the notational conventions of sect. 10.5. The strategy is to construct the full $\mathbf{T}$-matrix (34.3) from the exact 1-disk scattering matrix (35.9) 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 **T**-matrix in $S_{mm'} = \delta_{mm'} - i\,T_{mm'}$ can be split into a product over three kinds of matrices,

$$T_{mm'}(k) = \sum_{s,s'=1}^{N} \sum_{l_s,l_{s'}=-\infty}^{\infty} C_{ml_s}^{\;s}(k)\mathbf{M}^{-1}(k)^{ss'}_{l_s l_{s'}} D^{s'}_{l_{s'}m'}(k)\,.$$

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 $\Phi_s$ with respect to direction $\mathbf{k}$ of the outgoing spherical wave. As in (35.9), the matrix $\mathbf{C}^s$ takes form

$$C_{ml_s}^{\;s} = \frac{2i}{\pi a_s}\frac{J_{m-l_s}(kR_s)}{H^{(1)}_{l_s}(ka_s)}e^{im\Phi_s}\,. \tag{35.10}$$

If we now describe the ingoing spherical wave in the disk $s'$ coordinate frame by the matrix $\mathbf{D}^{s'}$

$$D^{s'}_{l_{s'}m'} = -\pi a_{s'} J_{m'-l_{s'}}(kR_{s'})J_{l_{s'}}(ka_{s'})e^{-im'\Phi_{s'}}\,, \tag{35.11}$$

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 **T**-matrix (35.9) for the single disk $s = s'$, $M = 1$ scattering. The Bessel function sum is a statement of the completness 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 **M** is called the *multi-scattering matrix*. If the scattering problem consists only of one scatterer, **M** is simply the unit matrix $M^{ss'}_{l_s l_{s'}} = \delta^{ss'}\delta_{l_s l_{s'}}$. For scattering from more than one scatterer we separate out a "single traversal" matrix **A** which transports the scattered wave from a scattering region $\mathcal{M}_s$ to the scattering region $\mathcal{M}_{s'}$,

$$M^{ss'}_{l_s l_{s'}} = \delta^{ss'}\delta_{l_s l_{s'}} - A^{ss'}_{l_s l_{s'}}\,. \tag{35.12}$$

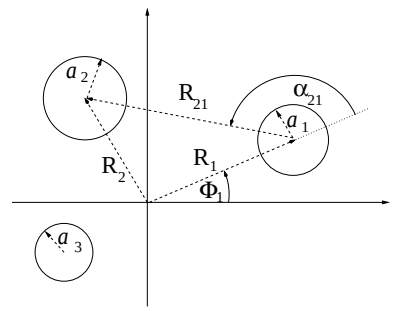



**Figure 35.1:** Global and local coordinates for a general 3-disk problem.

The matrix $\mathbf{A}^{ss'}$ reads:

$$A^{ss'}_{l_s l_{s'}} = -(1-\delta^{ss'})\,\frac{a_s}{a_{s'}}\,\frac{J_{l_s}(ka_s)}{H^{(1)}_{l_{s'}}(ka_{s'})}\,H^{(1)}_{l_s-l_{s'}}(kR_{ss'})\,e^{i(l_s\alpha_{s's}-l_{s'}(\alpha_{ss'}-\pi))}\,. \qquad (35.13)$$

Here, $a_s$ is the radius of the $s$th disk. $R_s$ and $\Phi_s$ 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 35.1).

Expanded as a geometrical series about the unit matrix **1**, the inverse matrix $\mathbf{M}^{-1}$ generates a multi-scattering series in powers of the single-traversal matrix **A**. All genuine multi-scattering dynamics is contained in the matrix **A**; by construction **A** vanishes for a single-scatterer system.

## 35.2 $N$-scatterer spectral determinant

In the following we limit ourselves to a study of the spectral properties of the **S**-matrix: resonances, time delays and phase shifts. The resonances are given by the poles of the **S**-matrix in the lower complex wave number ($k$) plane; more precisely, by the poles of the **S** on the second Riemann sheet of the complex energy plane. As the **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 **S**-matrix itself depends on the global coordinate system and on the choice of basis for the point particle wave function.

As the **S**-matrix is, in general, an infinite dimensional matrix, it is not clear whether the corresponding determinant exists at all. If **T**-matrix is trace-class, the determinant does exist. What does this mean?

### 35.2.1 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 J; here we will assume that the **T**-matrix (34.3) is trace-class, and draw the conlusions.

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

$$\det(\mathbf{1} - z\mathbf{A}) = \exp[\mathrm{tr}\ \ln(\mathbf{1} - z\mathbf{A})] = \exp\left(-\sum_{n=1}^{\infty}\frac{z^n}{z^n}\mathrm{tr}\,(\mathbf{A}^n)\right) .$$

That means

$$\det(\mathbf{1} - z\mathbf{A}) := \sum_{m=0}^{\infty} z^m Q_m(\mathbf{A}) , \tag{35.14}$$

where the cumulants $Q_m(\mathbf{A})$ satisfy the Plemelj-Smithies recursion formula (J.19), a generalization of Newton's formula to determinants of infinite-dimensional matrices,

$$\begin{aligned} Q_0(\mathbf{A}) &= 1 \\ Q_m(\mathbf{A}) &= -\frac{1}{m}\sum_{j=1}^{m} Q_{m-j}(\mathbf{A})\,\mathrm{tr}\,(\mathbf{A}^j) \quad \text{for } m \geq 1 , \end{aligned} \tag{35.15}$$

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 $\mathbf{C}^s$, $\mathbf{D}^s$ and $\mathbf{A}^{ss'}$? Manipulating the operators as though they were finite matrices, we can perform the following transformations:

$$\begin{aligned} \det\mathbf{S} &= \det\left(\mathbf{1} - i\mathbf{C}\mathbf{M}^{-1}\mathbf{D}\right) \\ &= \mathrm{Det}\left(\mathbf{1} - i\mathbf{M}^{-1}\mathbf{D}\mathbf{C}\right) = \mathrm{Det}\left(\mathbf{M}^{-1}(\mathbf{M} - i\mathbf{D}\mathbf{C})\right) \\ &= \frac{\mathrm{Det}\,(\mathbf{M} - i\mathbf{D}\mathbf{C})}{\mathrm{Det}\,(\mathbf{M})} \,. \end{aligned} \tag{35.16}$$

In the first line of (35.16) the determinant is taken over small $\ell$ (the angular momentum with respect to the global system). In the remainder of (35.16) 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 ... and tr ... refer to the $|\ell\rangle$ space, Det ... and Tr... 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 (35.16):

$$\begin{aligned}
\det \mathbf{S} &= \frac{\mathrm{Det}\,(\mathbf{M} - i\mathbf{DC})}{\mathrm{Det}\,(\mathbf{M})} \\
&= \frac{\mathrm{Det}\,(\mathbf{M} - i\mathbf{DC})}{\mathrm{Det}\,\mathbf{M}} \frac{\prod_{s=1}^{N} \det \mathbf{S}^s}{\prod_{s=1}^{N} \det \mathbf{S}^s} \\
&= \left(\prod_{s=1}^{N} \det \mathbf{S}^s\right) \frac{\mathrm{Det}\,(\mathbf{M} - i\mathbf{DC}) / \prod_{s=1}^{N} \det \mathbf{S}^s}{\mathrm{Det}\,\mathbf{M}} \,. \qquad (35.17)
\end{aligned}$$

The final step in the reformulation of the determinant of the **S**-matrix of the $N$-scatterer problem follows from the unitarity of the **S**-matrix. The unitarity of $\mathbf{S}^\dagger(k^*)$ implies for the determinant

$$\det\left(\mathbf{S}(k^*)^\dagger\right) = 1/\det \mathbf{S}(k)\,, \qquad (35.18)$$

where this manipulation is allowed because the **T**-matrix is trace-class. The unitarity condition should apply for the **S**-matrix of the total system, **S**, as for the each of the single subsystems, $\mathbf{S}^s$, $s = 1, \cdots, N$. In terms of the result of (35.17), this implies

$$\frac{\mathrm{Det}\,(\mathbf{M}(k) - i\mathbf{D}(k)\mathbf{C}(k))}{\prod_{s=1}^{N} \det \mathbf{S}^s} = \mathrm{Det}\,(\mathbf{M}(k^*)^\dagger)$$

since all determinants in (35.17) exist separately and since the determinants $\det \mathbf{S}^s$ respect unitarity by themselves. Thus, we finally have

$$\det \mathbf{S}(k) = \left\{\prod_{s=1}^{N} \left(\det \mathbf{S}^s(k)\right)\right\} \frac{\mathrm{Det}\,\mathbf{M}(k^*)^\dagger}{\mathrm{Det}\,\mathbf{M}(k)}\,, \qquad (35.19)$$

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 **T**-matrix belonging to

the total system and of the single-traversal matrix **A** and finally unitarity of the **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 insided 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 **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.

### 35.2.2 Quantum cycle expansions

In formula (35.19) 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 (35.19), 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 $\mathrm{Det}\,\mathbf{M}(k)$ 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 **A**, by means of the cumulant expansion (35.14). Because of the finite number $N \geq 2$ of scatterers $\mathrm{tr}\,(\mathbf{A}^n)$ 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 $s_i \in \{1, 2, \ldots, N\}$. of $N$ symbols,

$$\begin{aligned} &\mathrm{tr}\,\mathbf{A}^{s_1 s_2}\mathbf{A}^{s_2 s_3}\cdots\mathbf{A}^{s_{n-1}s_n}\mathbf{A}^{s_n s_n} \\ &= \sum_{l_{s_1}=-\infty}^{+\infty}\sum_{l_{s_2}=-\infty}^{+\infty}\cdots\sum_{l_{s_n}=-\infty}^{+\infty} A^{s_1 s_2}_{l_{s_1} l_{s_2}} A^{s_2 s_3}_{l_{s_2} l_{s_3}} \cdots A^{s_{n-1} s_n}_{l_{s_{n-1}} l_{s_n}} A^{s_n s_1}_{l_{s_n} l_{s_1}} \,. \end{aligned} \tag{35.20}$$

Remember our notation that the trace $\mathrm{tr}\,(\cdots)$ refers only to the $|l\rangle$ space. By construction **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 $1 - \delta^{ss'}$ in (35.13).

The trace $\mathrm{tr}\,\mathbf{A}^n$ is the sum of all itineraries of length $n$,

$$\mathrm{tr}\,\mathbf{A}^n = \sum_{\{s_1 s_2 \cdots s_n\}} \mathrm{tr}\,\mathbf{A}^{s_1 s_2}\mathbf{A}^{s_2 s_3}\cdots\mathbf{A}^{s_{n-1}s_n}\mathbf{A}^{s_n s_1} \,. \tag{35.21}$$

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. 35.3.1. Furthermore, depending on the geometry, there might be nontrivial pruning rules based on the so called ghost orbits, see sect.35.4.1.

### 35.2.3 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

$$\mathrm{Det}\,\mathbf{M} = \prod_{\alpha} (\det \mathbf{M}_{\mathrm{D}_\alpha}(k))^{d_\alpha} \quad ,$$

where the index $\alpha$ runs over all conjugate classes of the symmetry group $G$ and $\mathrm{D}_\alpha$ is the $\alpha$th representation of dimension $d_\alpha$. The symmetry reduction on the exact quantum mechanical level is the same as for the classical evolution operators spectral determinant factorization (19.17) of sect. 19.4.2.

## 35.3 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 (35.9), we do so for

$$\mathbf{d}(k) \equiv \frac{1}{2\pi i}\frac{d}{dk}\ln\det \mathbf{S}^1(ka) \;=\; \frac{1}{2\pi i}\frac{d}{dk}\mathrm{tr}\left(\ln \mathbf{S}^1(ka)\right) \tag{35.22}$$

the so called *time delay*.

$$\begin{aligned}\mathbf{d}(k) &= \frac{1}{2\pi i}\frac{d}{dk}\mathrm{tr}\left(\ln\det \mathbf{S}^1(ka)\right) = \frac{1}{2\pi i}\sum_m \left(\frac{H_m^{(1)}(ka)}{H_m^{(2)}(ka)}\frac{d}{dk}\frac{H_m^{(2)}(ka)}{H_m^{(1)}(ka)}\right)\\ &= \frac{a}{2\pi i}\sum_m \left(\frac{H_m^{(2)\prime}(ka)}{H_m^{(2)}(ka)} - \frac{H_m^{(1)\prime}(ka)}{H_m^{(1)}(ka)}\right) .\end{aligned} \tag{35.23}$$

Here the prime denotes the derivative with respect to the argument of the Hankel functions. Let us introduce the abbreviation

$$\chi_\nu = \frac{H_\nu^{(2)\prime}(ka)}{H_\nu^{(2)}(ka)} - \frac{H_\nu^{(1)\prime}(ka)}{H_\nu^{(1)}(ka)} \; . \tag{35.24}$$

We apply the Watson contour method to (35.23)

$$\mathbf{d}(k) = \frac{a_j}{2\pi i} \sum_{m=-\infty}^{+\infty} \chi_m = \frac{a_j}{2\pi i} \frac{1}{2i} \oint_C \mathrm{d}\nu \, \frac{e^{-i\nu\pi}}{\sin(\nu\pi)} \chi_\nu \, . \tag{35.25}$$

Here the contour $C$ encircles in a counter-clockwise manner a small semiinfinite strip $D$ which completely covers the real $\nu$-axis but which only has a small finite extent into the positive and negative imaginary $\nu$ direction. The contour $C$ is then split up in the path above and below the real $\nu$-axis such that

$$\mathbf{d}(k) = \frac{a}{4\pi i} \left\{ \int_{-\infty+i\epsilon}^{+\infty+i\epsilon} \mathrm{d}\nu \, \frac{e^{-i\nu\pi}}{\sin(\nu\pi)} \chi_\nu - \int_{-\infty-i\epsilon}^{+\infty-i\epsilon} \mathrm{d}\nu \, \frac{e^{-i\nu\pi}}{\sin(\nu\pi)} \chi_\nu \right\} \, .$$

Then, we perform the substitution $\nu \to -\nu$ in the second integral so as to get

$$\begin{aligned} \mathbf{d}(k) &= \frac{a}{4\pi} \left\{ \int_{-\infty+i\epsilon}^{+\infty+i\epsilon} \mathrm{d}\nu \, \frac{e^{-i\nu\pi}}{\sin(\nu\pi)} \chi_\nu + \mathrm{d}\nu \, \frac{e^{+i\nu\pi}}{\sin(\nu\pi)} \chi_{-\nu} \right\} \\ &= \frac{a}{2\pi i} \left\{ 2 \int_{-\infty+i\epsilon}^{+\infty+i\epsilon} \mathrm{d}\nu \, \frac{e^{2i\nu\pi}}{1-e^{2i\nu\pi}} \chi_\nu + \int_{-\infty}^{+\infty} \mathrm{d}\nu \, \chi_\nu \right\} \, , \end{aligned} \tag{35.26}$$

where we used the fact that $\chi_{-\nu} = \chi_\nu$. The contour in the last integral can be deformed to pass over the real $\nu$-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 (35.26) 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 $\chi_\nu$, the second integral will be evaluated on the real $\nu$-axis under the Debye approximation for Hankel functions.

We will now work out the first term. The poles of $\chi_\nu$ in the upper complex plane are given by the zeros of $H_\nu^{(1)}(ka)$ which will be denoted by $\nu_\ell(ka)$ and by the zeros of $H_\nu^{(2)}(ka)$ which we will denote by $-\bar{\nu}_\ell(ka)$, $\ell = 1, 2, 3, \cdots$. In the Airy approximation to the Hankel functions they are given by

$$\nu_\ell(ka) = ka + i\alpha_\ell(ka) \, , \tag{35.27}$$

$$-\bar{\nu}_\ell(ka) = -ka + i(\alpha_\ell(k^*a))^* = -\left(\nu_\ell(k^*a)\right)^* \, , \tag{35.28}$$

with

$$\begin{aligned} i\alpha_\ell(ka) &= e^{i\frac{\pi}{3}} \left(\frac{ka}{6}\right)^{1/3} q_\ell - e^{-i\frac{\pi}{3}} \left(\frac{6}{ka}\right)^{1/3} \frac{q_\ell^2}{180} - \frac{1}{70ka} \left(1 - \frac{q_\ell^3}{30}\right) \\ &+ e^{i\frac{\pi}{3}} \left(\frac{6}{ka}\right)^{\frac{5}{3}} \frac{1}{3150} \left(\frac{29 q_\ell}{6^2} - \frac{281 q_\ell^4}{180 \cdot 6^3}\right) + \cdots . \end{aligned} \tag{35.29}$$

Here $q_\ell$ labels the zeros of the Airy integral

$$A(q) \equiv \int_0^\infty d\tau\, \cos(q\tau - \tau^3) = 3^{-1/3}\pi \mathrm{Ai}(-3^{-1/3}q)\ ,$$

with 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 (35.26) 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 (35.26) we will insert the Debye approximations for the Hankel functions:

$$\begin{aligned} H_\nu^{(1/2)}(x) &\sim \sqrt{\frac{2}{\pi\sqrt{x^2-\nu^2}}}\exp\left(\pm i\sqrt{x^2-\nu^2} \mp i\nu \arccos\frac{\nu}{x} \mp i\frac{\pi}{4}\right) \quad \text{for } |x| > \nu \\ H_\nu^{(1/2)}(x) &\sim \mp i\sqrt{\frac{2}{\pi\sqrt{\nu^2-x^2}}}\exp\left(-\sqrt{\nu^2-x^2} + \nu \mathrm{ArcCosh}\frac{\nu}{x}\right) \quad \text{for } |x| < \nu\ . \end{aligned} \tag{35.30}$$

Note that for $\nu > ka$ the contributions in $\chi_\nu$ cancel. Thus the second integral of (35.26) becomes

$$\begin{aligned} \frac{a}{2\pi i}\int_{-\infty}^{+\infty} d\nu\, \chi_\nu &= \frac{a}{2\pi i}\int_{-ka}^{+ka} d\nu\, \frac{(-2i)}{a}\frac{d}{dk}\left(\sqrt{k^2a^2-\nu^2} - \nu\arccos\frac{\nu}{ka}\right) + \cdots \\ &= -\frac{1}{k\pi}\int_{-ka}^{ka} d\nu\, \sqrt{k^2a^2-\nu^2} + \cdots = -\frac{a^2}{2}k + \cdots\ , \end{aligned} \tag{35.31}$$

where $\cdots$ takes care of the polynomial corrections in the Debye approximation and the boundary correction terms in the $\nu$ integration.

In summary, the semiclassical approximation to $\mathbf{d}(k)$ reads

$$\mathbf{d}(k) = 2a\sum_{\ell=1}^{\infty}\left(\frac{e^{2i\nu_\ell\pi}}{1-e^{2i\nu_\ell\pi}} + \frac{e^{-2i\bar{\nu}_\ell\pi}}{1-e^{-2i\bar{\nu}_\ell\pi}}\right) - \frac{a^2}{2}k + \cdots\ .$$

Using the definition of the time delay (35.22), we get the following expression for $\det \mathbf{S}^1(ka)$:

$$\begin{aligned} &\ln\det\mathbf{S}^1(ka) - \lim_{k_0\to 0}\ln\det\mathbf{S}^1(k_0a) \qquad (35.32) \\ &= 2\pi i a\int_0^k d\tilde{k}\left(-\frac{a\tilde{k}}{2} + 2\sum_{\ell=1}^{\infty}\left(\frac{e^{i2\pi\nu_\ell(\tilde{k}a)}}{1-e^{i2\pi\nu_\ell(\tilde{k}a)}} + \frac{e^{-i2\pi\bar{\nu}_\ell(\tilde{k}a)}}{1-e^{-i2\pi\bar{\nu}_\ell(\tilde{k}a)}}\right)\right) + \cdots \\ &\sim -2\pi i N(k) + 2\sum_{\ell=1}^{\infty}\int_0^k d\tilde{k}\,\frac{d}{d\tilde{k}}\left\{-\ln\left(1-e^{i2\pi\nu_\ell(\tilde{k}a)}\right) + \ln\left(1-e^{-i2\pi\bar{\nu}_\ell(\tilde{k}a)}\right)\right\} + \cdots, \end{aligned}$$

where in the last expression it has been used that semiclassically $\frac{d}{dk}\nu_\ell(ka) \sim \frac{d}{dk}\bar{\nu}_\ell(ka) \sim a$ and that the Weyl term for a single disk of radius $a$ goes like $N(k) = \pi a^2 k^2/(4\pi) + \cdots$ (the next terms come from the boundary terms in the $\nu$-integration in (35.31)). Note that for the lower limit, $k_0 \to 0$, we have two simplifications: First,

$$\begin{aligned}\lim_{k_0\to 0} S^1_{mm'}(k_0 a) = \lim_{k_0\to 0} \frac{-H_m^{(2)}(k_0 a)}{H_m^{(1)}(k_0 a)}\delta_{mm'} &= 1\times\delta_{mm'} \qquad \forall m, m' \\ \to \quad \lim_{k_0\to 0} \det \mathbf{S}^1(k_0 a) &= 1\ .\end{aligned}$$

Secondly, for $k_0 \to 0$, the two terms in the curly bracket of (35.32) cancel.

### 35.3.1 1-disk spectrum interpreted; pure creeping

To summarize: the semiclassical approximation to the determinant $\mathbf{S}^1(ka)$ is given by

$$\det \mathbf{S}^1(ka) \sim e^{-i2\pi N(k)} \frac{\prod_{\ell=1}^{\infty}\left(1 - e^{-2i\pi\bar{\nu}_\ell(ka)}\right)^2}{\prod_{\ell=1}^{\infty}\left(1 - e^{2i\pi\nu_\ell(ka)}\right)^2}\ , \tag{35.33}$$

with

$$\begin{aligned}\nu_\ell(ka) &= ka + i\alpha_\ell(ka) &&= ka + e^{+i\pi/3}(ka/6)^{1/3}q_\ell + \cdots \\ \bar{\nu}_\ell(ka) &= ka - i(\alpha_\ell(k^*a))^* &&= ka + e^{-i\pi/3}(ka/6)^{1/3}q_\ell + \cdots \\ &= (\nu_\ell(k^*a))^*\end{aligned}$$

and $N(ka) = (\pi a^2 k^2)/4\pi + \cdots$ 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 (35.33) for a disk $s$ as

$$\det \mathbf{S}^s(ka_s) \sim \left(e^{-i\pi N(ka_s)}\right)^2 \frac{\widetilde{Z}^s_\ell(k^*a_s)^*}{\widetilde{Z}^s_\ell(ka_s)} \frac{\widetilde{Z}^s_r(k^*a_s)^*}{\widetilde{Z}^s_r(ka_s)}\ , \tag{35.34}$$

where $\widetilde{Z}^s_\ell(ka_s)$ and $\widetilde{Z}^s_r(ka_s)$ are the *diffractional* 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 35.2). The two orientations of the creeping

**Figure 35.2:** Right- and left-handed diffractive creeping paths of increasing mode number $\ell$ for a single disk.

orbits are the reason for the exponents 2 in (35.33). Equation (35.33) describes the semiclassical approximation to the incoherent part (= the curly bracket on the r.h.s.) of the exact expression (35.19) 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 (35.9). The exact quantummechanical poles of the scattering matrix are therefore given by the the zeros, $k^{\rm res}_{n\,m}$, of the Hankel functions $H^{(1)}_m(ka)$ 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 $H^{(1)}_\nu(ka)$ of complex-valued index $\nu$ reads

$$H^{(1)}_\nu(ka) \sim \frac{2}{\pi} e^{-i\frac{\pi}{3}} \left(\frac{6}{ka}\right)^{1/3} A(q^{(1)}) \ ,$$

with

$$q^{(1)} = e^{-i\frac{\pi}{3}} \left(\frac{6}{ka}\right)^{1/3} (\nu - ka) + O\left((ka)^{-1}\right) .$$

Hence the zeros $\nu_\ell$ of the Hankel function in the complex $\nu$ plane follow from the zeros $q_\ell$ of the Airy integral $A(q)$ (see (35.3). Thus if we set $\nu_\ell = m$ (with $m$ integer), we have the following semiclassical condition on $k^{\rm res}$

$$\begin{aligned} m &\sim k^{\rm res}a + i\alpha_\ell(k^{\rm res}a) \\ &= e^{i\frac{\pi}{3}} \left(\frac{k^{\rm res}a}{6}\right)^{1/3} q_\ell \ - \ e^{-i\frac{\pi}{3}} \left(\frac{6}{k^{\rm res}a}\right)^{1/3} \frac{q_\ell^2}{180} \ - \ \frac{1}{70k^{\rm res}a}\left(1 - \frac{q_\ell^3}{30}\right) \end{aligned}$$

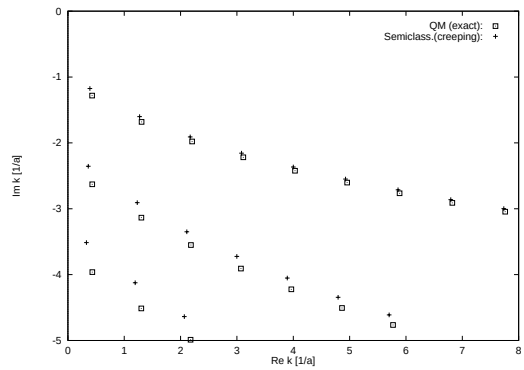



**Figure 35.3:** The shape resonances of the 1-disk system in the complex $k$ plane in units of the disk radius $a$. The boxes label the exact quantum mechanical resonances (given by the zeros of $H_m^{(1)}(ka)$ for $m = 0, 1, 2$), the crosses label the diffractional semiclassical resonances (given by the zeros of the creeping formula in the Airy approximation (35.35) up to the order $O([ka]^{1/3})$).

$$+ \quad e^{i\frac{\pi}{3}} \left(\frac{6}{k^{\rm res}a}\right)^{\frac{1}{3}} \frac{1}{3150} \left(\frac{29q_\ell}{6^2} - \frac{281q_\ell^4}{180 \cdot 6^3}\right) + \cdots,$$

$$\text{with } l = 1, 2, 3, \cdots. \tag{35.35}$$

For a given index $l$ this is equivalent to

$$0 \sim 1 - e^{(ik^{\rm res} - \alpha_\ell)2\pi a},$$

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

$$\prod_{l=1}^{\infty} \left(1 - e^{(ik - \alpha_\ell)2\pi a}\right),$$

which is of course nothing else than $\widetilde{Z}_{1\text{-disk}}(k)$, the semiclassical diffraction zeta function of the 1-disk scattering problem, see (35.34). 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.

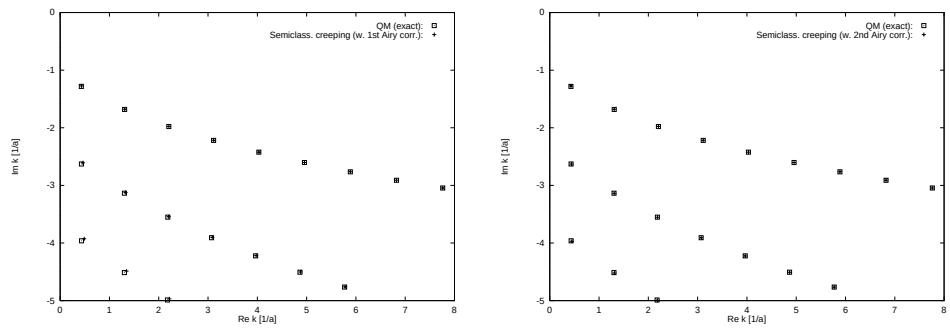



**Figure 35.4:** Same as in figure 35.3. However, the subleading terms in the Airy approximation (35.35) are taken into account up to the order $O([ka]^{-1/3})$ (upper panel) and up to order $O([ka]^{-1})$ (lower panel).

From figure 35.3 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 $\mathrm{Re}\,k$ and decreasing $|\mathrm{Im}\,k|$, as they should as semiclassical approximations.

In the upper panel of figure 35.4 one sees the change, when the next order in the Airy approximation (35.35) is taken into account. The approximation is nearly perfect, especially for the leading row of resonances. The second Airy approximation using (35.35) up to order $O([ka]^{-1})$ is perfect up to the drawing scale of figure 35.4 (lower panel).

## 35.4 From quantum cycle to semiclassical cycle

The procedure for the semiclassical approximation of a general periodic itinerary (35.20) of length $n$ 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

$$\mathrm{tr}\,\mathbf{A}^{s_1 s_2}\cdots\mathbf{A}^{s_m s_1} = \sum_{l_{s_1}=-\infty}^{\infty}\cdots\sum_{l_{s_m}=-\infty}^{\infty} A^{s_1 s_2}_{l_{s_1} l_{s_2}}\cdots A^{s_m s_1}_{l_{s_m} l_{s_1}}$$

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 $\nu_{s_i}$. Then the paths below the real $\nu_{s_i}$-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.

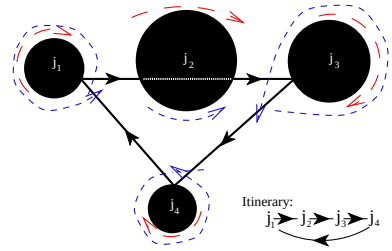



**Figure 35.5:** A 4-disk problem with three specular reflections, one ghost tunneling, and distinct creeping segments from which all associated creeping paths can be constructed.

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 $\nu_{s_i}$ plane and evaluate the integral at the residua $H^{(1)}_{\nu_{s_i}}(ka_{s_i})$=0. Then we use the Airy approximation for $J_{\nu_{s_i}}(ka_{s_i})$ and $H^{(1)}_{\nu_{s_i}}(ka_{s_i})$: 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 coninciding 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 35.5. 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

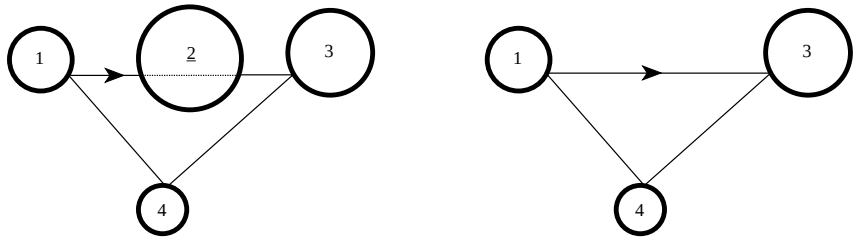



**Figure 35.6:** (a) The ghost itinerary $(1, \underline{2}, 3, 4)$. (b) The parent itinerary $(1, 3, 4)$.

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

$$\mathrm{tr}\, \mathbf{A}^{s_1 s_2} \mathbf{A}^{s_2 s_3} \cdots \mathbf{A}^{s_{n-1} s_n} \mathbf{A}^{s_n s_1}$$

becomes a standard periodic orbit labeled by the symbol sequence $s_1 s_2 \cdots s_n$. Depending on the geometry, the individual legs $s_{i-1} \to s_i \to s_{i+1}$ 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 $\ell_i$ 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.

### 35.4.1 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. Thus, semiclassically, they cancel each other in the $\mathrm{tr}\, \ln(\mathbf{1} - \mathbf{A})$ expansion, where they are multiplied by the permutation factor $n/r$ with the integer $r$ counting the repeats. For example, let $(1, \underline{2}, 3, 4)$ be a non-repeated periodic itinerary with a ghost section at disk 2 steming from the 4th-order trace $\mathrm{tr}\, A^4$. By convention, an underlined disk index signals a ghost passage (as in figure 35.6a), with corresponding semiclassical ghost traversal matrices also underlined, $\underline{\mathbf{A}}^{i,\underline{i+1}} \underline{\mathbf{A}}^{\underline{i+1},i+2}$. Then its semiclassical, geometrical contribution to $\mathrm{tr}\, \ln(\mathbf{1} - \mathbf{A})$ cancels exactly against the one of its "parent" itinerary $(1, 3, 4)$ (see figure 35.6b) resulting from the 3rd-order trace:

$$-\frac{1}{4}\left(4\, \underline{\mathbf{A}}^{1,\underline{2}} \underline{\mathbf{A}}^{\underline{2},3} \mathbf{A}^{3,4} \mathbf{A}^{4,1}\right) - \frac{1}{3}\left(3\, \mathbf{A}^{1,3} \mathbf{A}^{3,4} \mathbf{A}^{4,1}\right)$$

[1] Actually, these are double-letter symbols as $\sigma_i$ and $l_i$ are only counted as a product.

$$= \quad (+1-1)\,\mathbf{A}^{1,3}\mathbf{A}^{3,4}\mathbf{A}^{4,1} = 0\ .$$

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

$$\cdots \underline{\mathbf{A}}^{i,\underline{i+1}}\underline{\mathbf{A}}^{\underline{i+1},i+2}\cdots = \cdots\left(-\mathbf{A}^{i,i+2}\right)\cdots\ . \tag{35.36}$$

The reader might study more complicated examples and convince herself that the rule (35.36).is sufficient to cancel any primary or repeated periodic orbit with one or more ghost sections completely out of the expansion of tr $\ln(\mathbf{1}-\mathbf{A})$ and therefore also out of the cumulant expansion in the semiclassical limit: Any periodic orbit of length $m$ with $n(<m)$ ghost sections is cancelled by the sum of all 'parent' periodic orbits of length $m-i$ (with $1 \le i \le n$ 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, tr $\mathbf{A}^2$, as tr $\mathbf{A}$ itself vanishes identically in the full domain which is considered here.

## 35.5 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 classical 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 35.1** Sources. This chapter is based in its entirety on ref. [1]; the reader is referred to the full exposition for the proofs and discussion of details omitted here.

sect. 35.3 is based on appendix E of ref. [1]. We follow Franz [19] in applying the Watson contour method [20] to (35.23). The Airy and Debye approximations to the Hankel functions are given in ref. [21], the Airy expansion of the 1-disk zeros can be found in ref. [22].For details see refs. [19, 22, 23, 1]. That the interior domains of the disks are excluded relatively to the free evolution without scattering obstacles was noted in refs. [24, 15].

The procedure for the semiclassical approximation of a general periodic itinerary (35.20) of length $n$ can be found in ref. [1] 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. [12, 24].

**Remark 35.2** Krein-Friedel-Lloyd formula. In the literature (see, e.g., refs. [14, 15] based on ref. [11] or ref. [1]) 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 [5, 6, 8, 9]. See also ref. [13] for a modern discussion of the Krein-Friedel-Lloyd formula and refs. [1, 17] for the connection of (34.17) to the the Wigner time delay.

The order of the two limits in (34.18) and (34.17) is essential, see e.g. Balian and Bloch [11] 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 $b \to \infty$, if they were not damped out by the $+i\epsilon$ term.

[exercise 34.1]

**Remark 35.3** $\mathbf{T}$, $\mathbf{C}^s$, $\mathbf{D}^s$ and $\mathbf{A}^{ss'}$ matrices are trace-class In refs. [1] 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 particulary easy to prove.

# Chapter 36

# 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 classical 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 classical 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 trajectories in sect. 6.3. Here we will find periodic orbits and determine the relevant eigenvalues of the fundamental matrix in sect. 36.1. We will explain in

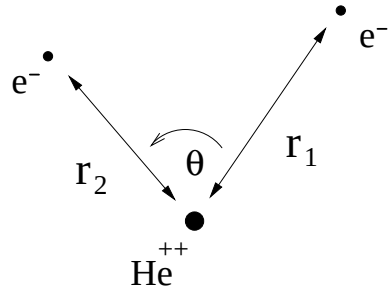



**Figure 36.1:** Coordinates for the helium three body problem in the plane.

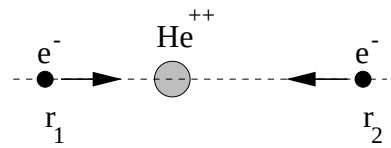



**Figure 36.2:** Collinear helium, with the two electrons on opposite sides of the nucleus.

sect. 36.5 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. 36.5.1.

## 36.1 Classical dynamics of collinear helium

Recapitulating briefly what we learned in sect. 6.3: 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 36.2.

The Hamiltonian can be brought to a non–dimensionalized form

$$H = \frac{p_1^2}{2} + \frac{p_2^2}{2} - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_1 + r_2} = -1\,. \tag{36.1}$$

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. 36.5.

There is another classical 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 classical action (32.15) which scales with energy as

$$S(E) = \oint \mathrm{d}\mathbf{q}(E)\cdot\mathbf{p}(E) = \frac{e^2 m_e^{1/2}}{(-E)^{1/2}}\, S\,, \tag{36.2}$$

with $S$ being the action obtained from (36.1) for $E = -1$, and coordinates $\mathbf{q} = (r_1, r_2)$, $\mathbf{p} = (p_1, p_2)$. For the Hamiltonian (36.1), the period of a cycle and its action are related by (32.17), $T_p = \frac{1}{2} S_p$.

After a Kustaanheimo–Stiefel transformation

$$r_1 = Q_1^2\,, \qquad r_2 = Q_2^2\,, \qquad p_1 = \frac{P_1}{2Q_1}\,, \qquad p_2 = \frac{P_2}{2Q_2}\,, \tag{36.3}$$

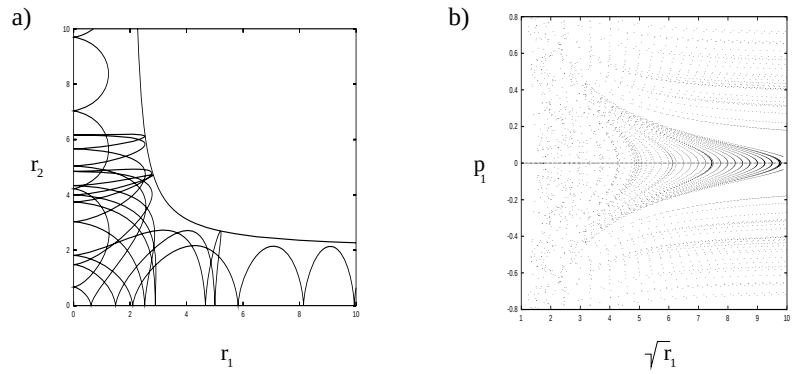



**Figure 36.3:** (a) A typical trajectory in the $r_1$ – $r_2$ plane; the trajectory enters here along the $r_1$ axis and escapes to infinity along the $r_2$ axis; (b) Poincaré map ($r_2$=0) for collinear helium. Strong chaos prevails for small $r_1$ near the nucleus.

and reparametrization of time by $d\tau = dt/r_1 r_2$, the equations of motion take form (6.19)

[exercise 36.1]

$$\dot{P}_1 = 2Q_1\left[2 - \frac{P_2^2}{8} - Q_2^2\left(1 + \frac{Q_2^2}{R_{12}^4}\right)\right]; \qquad \dot{Q}_1 = \frac{1}{4}P_1 Q_2^2 \tag{36.4}$$
$$\dot{P}_2 = 2Q_2\left[2 - \frac{P_1^2}{8} - Q_1^2\left(1 + \frac{Q_1^2}{R_{12}^4}\right)\right]; \qquad \dot{Q}_2 = \frac{1}{4}P_2 Q_1^2.$$

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 (6.19) 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 (6.18) 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.

## 36.2 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.

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 36.3 (a) where we used $r_1$ and $r_2$ as the relevant axis. The dynamics can also be visualized in a Poincaré surface of section, see figure 36.3 (b). We plot here the coordinate and momentum of the outer electron whenever the inner particle hits

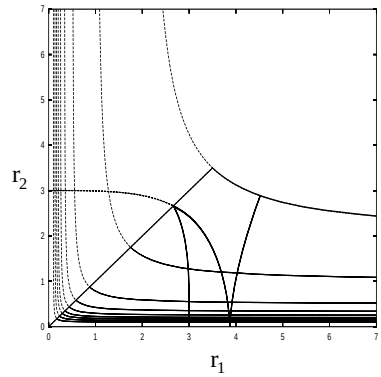



**Figure 36.4:** The cycle 011 in the fundamental domain $r_1 \geq r_2$ (full line) and in the full domain (dashed line).

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 fugative. 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.

### 36.2.1 Reflection symmetry

The Hamiltonian (6.10) is invariant with respect to electron–electron exchange; this symmetry corresponds to the mirror symmetry of the potential along the line $r_1 = r_2$, figure 36.4. 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 19, the dynamics in the fundamental domain is the key to the factorization of spectral determinants, to be implemented here in (36.15). Note also the similarity between the fundamental domain of the collinear potential figure 36.4, and the fundamental domain figure **??** (b) in the 3–disk system, a simpler problem with the same binary symbolic dynamics.

in depth:
sect. 19.6, p. 331

### 36.2.2 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 (36.1) 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 Poincaré 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 Poincaré 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 $\overline{0}$ cycle.

### 36.2.3 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. 13.5.2. 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 surface of section, e.g. $r_2 = 0$ or $r_1 = r_2$, leaves us in general with a 2-dimensional search. Methods to find periodic orbits in multidimensional spaces have been described in chapter 12. 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 based on the binary symbolic dynamics phase space partitioning. All cycles up to symbol length 16 (some 8000 primitive cycles) have been computed by such methods, with some examples shown in figure 36.5. All numerical evidence indicates that the dynamics of collinear helium is hyperbolic, and that all periodic orbits are unstable.

Note that the fixed point $\overline{0}$ cycle is not in this list. The $\overline{0}$ 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

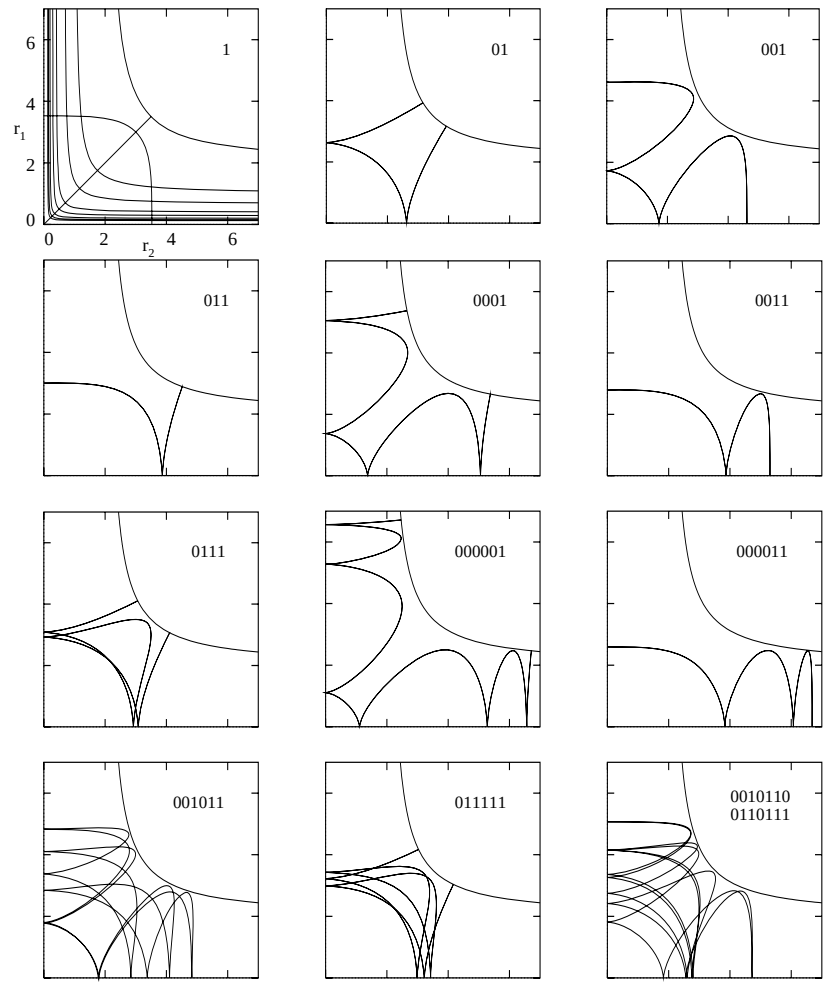



**Figure 36.5:** Some of the shortest cycles in collinear helium. The classical collinear electron motion is bounded by the potential barrier $-1 = -2/r_1 - 2/r_2 + 1/(r_1 + r_2)$ and the condition $r_i \geq 0$. The orbits are shown in the full $r_1$–$r_2$ domain, the itineraries refers to the dynamics 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.

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 36.3 (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 $(p_1, p_2) \to (-p_1, -p_2)$, 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 36.5 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 36.5;
- 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 36.5;
- 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 36.5.

All cycles up to symbol length 5 are time reversal invariant, the first two non-time reversal symmetric cycles are cycles $\overline{001011}$ and $\overline{001101}$ in figure 36.5. Their determination would require 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 classical collinear helium with the help of the equations of motions (6.19) and to find all cycles up to length 5. There is only one thing not yet in place; we need the governing equations for the matrix elements of the fundamental 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 B.4. [exercise 36.5]

## 36.3 Local coordinates, fundamental matrix

In this section, we will derive the equations of motion for the fundamental matrix along a collinear helium trajectory. The fundamental 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 B. We will give the transformation to local coordinates explicitly, here for the regularized coordinates (6.17), and state the resulting equations of motion for the reduced $[2 \times 2]$ fundamental matrix.

The vector locally parallel to the trajectory is pointing in the direction of the phase space velocity (7.7)

$$v_m = \dot{x}_m(t) = \omega_{mn} \frac{\partial H}{\partial x_n} = (H_{P_1}, H_{P_2}, -H_{Q_1}, -H_{Q_2})^T ,$$

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 (6.18)

$$\gamma = \nabla H = (H_{Q_1}, H_{Q_2}, H_{P_1}, H_{P_2})^T .$$

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

$$\begin{aligned} \mathbf{O} &= (\gamma_1, \gamma_2, \gamma/R, v) \qquad (36.5) \\ &= \begin{pmatrix} -H_{P_2}/R & H_{Q_2} & H_{Q_1}/R & H_{P_1} \\ H_{P_1}/R & -H_{Q_1} & H_{Q_2}/R & H_{P_2} \\ -H_{Q_2}/R & -H_{P_2} & H_{P_1}/R & -H_{Q_1} \\ H_{Q_1}/R & H_{P_1} & H_{P_2}/R & -H_{Q_2} \end{pmatrix} \end{aligned}$$

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 $(\gamma, v)$. 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. [exercise 36.6] The fundamental matrix (4.6) rotated to the local coordinate system by $\mathbf{O}$ then has the form

$$\mathbf{m} = \begin{pmatrix} m_{11} & m_{12} & * & 0 \\ m_{21} & m_{22} & * & 0 \\ 0 & 0 & 1 & 0 \\ * & * & * & 1 \end{pmatrix}, \quad M = \mathbf{O}^{\mathbf{T}} \mathbf{m} \mathbf{O}$$

The linearized motion perpendicular to the trajectory on the energy manifold is described by the $[2 \times 2]$ matrix $\mathbf{m}$; the 'trivial' directions correspond to unit eigenvalues on the diagonal in the 3rd and 4th column and row.

The equations of motion for the reduced fundamental matrix $\mathbf{m}$ are given by

$$\dot{\mathbf{m}} = \mathbf{l}(t)\mathbf{m}(t), \qquad (36.6)$$

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

$$l_{11} = \frac{1}{R}[2H_{Q_1Q_2}(H_{Q_2}H_{P_1} + H_{Q_1}H_{P_2}) \qquad (36.7)$$

| p | $S_p/2\pi$ | $\ln\lvert\Lambda_p\rvert$ | $\sigma_p$ | $m_p$ |
|---|---|---|---|---|
| 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 |

**Table 36.1:** Action $S_p$ (in units of $2\pi$), Lyapunov exponent $|\Lambda_p|/T_p$ for the motion in the collinear plane, winding number $\sigma_p$ for the motion perpendicular to the collinear plane, and the topological index $m_p$ for all fundamental domain cycles up to topological length 6.

$$
\begin{aligned}
& +(H_{Q_1}H_{P_1} - H_{Q_2}H_{P_2})(H_{Q_1Q_1} - H_{Q_2Q_2} - H_{P_1P_1} + H_{P_2P_2})] \\
l_{12} &= -2H_{Q_1Q_2}(H_{Q_1}H_{Q_2} - H_{P_1}H_{P_2}) \\
& +(H_{Q_1}^2 + H_{P_2}^2)(H_{Q_2Q_2} + H_{P_1P_1}) + (H_{Q_2}^2 + H_{P_1}^2)(H_{Q_1Q_1} + H_{P_2P_2}) \\
l_{21} &= \frac{1}{R^2}[2(H_{Q_1P_2} + H_{Q_2P_1})(H_{Q_2}H_{P_1} + H_{Q_1}H_{P_8}) \\
& -(H_{P_1}^2 + H_{P_2}^2)(H_{Q_1Q_1} + H_{Q_2Q_2}) - (H_{Q_1}^2 + H_{Q_2}^2)(H_{P_1P_1} + H_{P_2P_2})] \\
l_{22} &= -l_{11}\,.
\end{aligned}
$$

Here $H_{Q_iQ_j}$, $H_{P_iP_j}$, $i,\ j = 1,2$ are the second partial derivatives of $H$ with respect to the coordinates $Q_i$, $P_i$, evaluated at the phase space coordinate of the classical trajectory.

## 36.4 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 fundamental matrix whose eigenvalues enter the semiclassical spectral determinant (33.12). By (32.17) 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 fundamental matrix does pick up extra contributions. When we calculate the fundamental matrix for the full problem, we must also allow for displacements out of the collinear plane, so the full fundamental 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 fundamental matrix can be written in terms of two distinct $[2 \times 2]$ matrices, with trivial eigendirections 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 $\Theta$-coordinate in figure 36.1. It turns out that the linearized dynamics in the $\Theta$ 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. 36.5.

The numerical values of the actions, Floquet exponents, stability angles, and topological indices for the shortest cycles are listed in table **??**. These numbers, needed for the semiclassical quantization implemented in the next section, an also be helpful in checking your own calculations.

## 36.5 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 36.4.

### 36.5.1 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

$$E_N = -\frac{e^4 m_e}{\hbar^2}\frac{Z^2}{2N^2}\,, \tag{36.8}$$

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

$$E_{N,n} = -\frac{2}{N^2} - \frac{1}{2n^2} \qquad \text{with} \quad n > N. \tag{36.9}$$

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 $n \to \infty$, 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 classical analysis of the previous section in mind: we already found that one of the classical 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 (36.8) 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 $l = 0, 1 \ldots N - 1$. 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 classical 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 (36.9) 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 36.6. The 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 classical 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 36.6. We

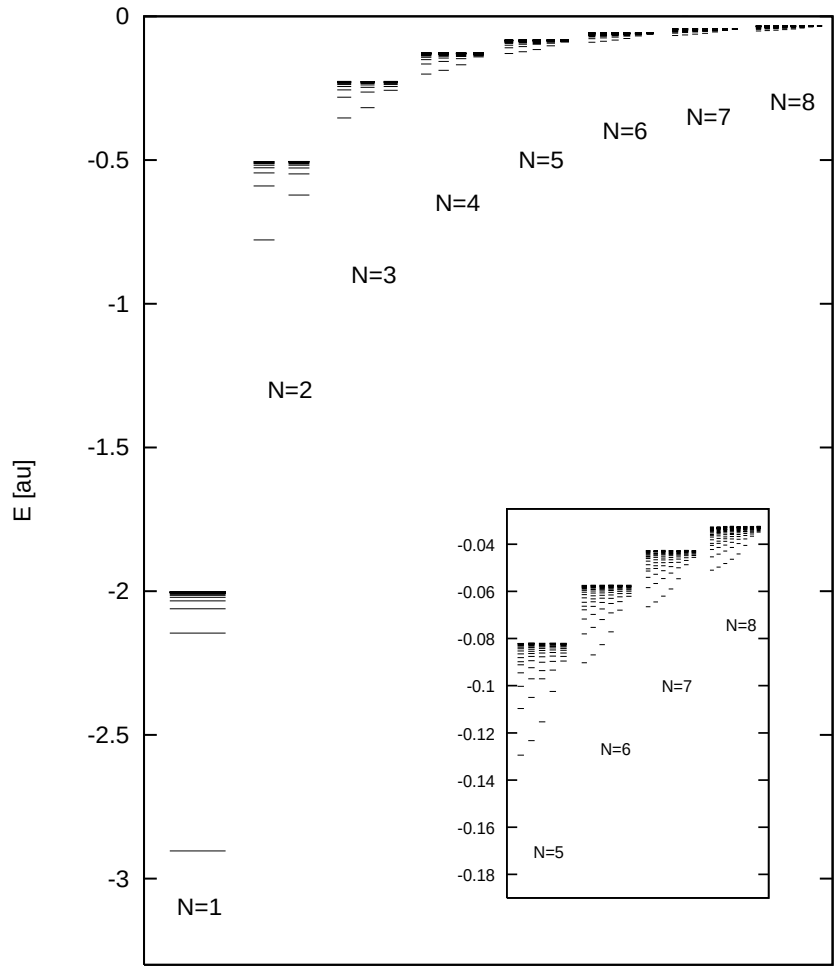



**Figure 36.6:** The exact quantum helium spectrum for $L = 0$. The energy levels denoted by bars have been obtained from full 3-dimensional quantum calculations [3].

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 moderate angular momentum $l$ as long as $l \ll N$. . [remark 36.4]

### 36.5.2 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 (33.12) has been written as product over all cycles of the classical systems. The energy dependence in collinear helium enters the classical dynamics only through simple scaling transformations described in sect. 6.3.1 which makes it possible to write the semiclassical spectral determinant in the form

$$\det(\hat{H}-E)_{sc} = \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{ir(sS_p - m_p\frac{\pi}{2})}}{(-\det(1-M^r_{p\perp}))^{1/2}|\det(1-M^r_{p\|})|^{1/2}}\right), \quad (36.10)$$

with the energy dependence absorbed into the variable

$$s = \frac{e^2}{\hbar}\sqrt{\frac{m_e}{-E}}\,,$$

obtained by using the scaling relation (36.2) for the action. As explained in sect. 36.3, the fact that the $[4\times 4]$ fundamental 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, 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 $e^{i\pi\bar{N}(E)}$ present in (33.12) is absent in (36.10). 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 $\bar{N}(E)$ not defined. In order to obtain a spectral determinant as an infinite product of the form (33.18) we may proceed as in (17.9) by expanding the determinants in (36.10) in terms of the eigenvalues of the corresponding fundamental 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. $\sigma$ 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 (17.9) and (33.18) we may thus write

$$\begin{aligned}
&\left[-\det(1-M_\perp^r)|\det(1-M_\parallel^r)|\right]^{-1/2} \qquad (36.11)\\
&= \left[-(1-\Lambda^r)(1-\Lambda^{-r})|(1-e^{2\pi i r\sigma})(1-e^{-2\pi i r\sigma})\right]^{-1/2}\\
&= \sum_{k,\ell=0}^{\infty}\frac{1}{|\Lambda^r|^{1/2}\Lambda^{rk}}e^{-ir(\ell+1/2)\sigma}\,.
\end{aligned}$$

The $\pm$ sign corresponds to the hyperbolic/inverse hyperbolic periodic orbits with positive/negative eigenvalues $\Lambda$. Using the relation (36.12) we see that the sum over $r$ in (36.10) is the expansion of the logarithm, so the semiclassical spectral determinant can be rewritten as a product over dynamical zeta functions, as in (17.9):

$$\det(\hat{H}-E)_{sc} = \prod_{k=0}^{\infty}\prod_{m=0}^{\infty}\zeta_{k,m}^{-1} = \prod_{k=0}^{\infty}\prod_{m=0}^{\infty}\prod_{p}(1-t_p^{(k,m)})\,, \qquad (36.12)$$

where the cycle weights are given by

$$t_p^{(k,m)} = \frac{1}{|\Lambda|^{1/2}\Lambda^k}e^{i\left(sS_p - m_p\frac{\pi}{2} - 4\pi(\ell+1/2)\sigma_p\right)}\,, \qquad (36.13)$$

and $m_p$ is the topological index for the motion in the collinear plane which equals twice the topological length 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 $\sigma$. The values for the actions, winding numbers and stability indices of the shortest cycles in collinear helium are listed in table **??**.

The integer indices $\ell$ and $k$ play very different roles in the semiclassical spectral determinant (36.12). 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 $\ell$ 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 $\Theta$ coordinate and can in our simplified picture developed in sect. 36.5.1 be related to the quantum number $l = l_1 = l_2$ representing the single particle angular momenta. Every distinct $\ell$ value corresponds to a full spectrum which we obtain from the zeros of the semiclassical spectral determinant $1/\zeta_\ell$ keeping $\ell$ fixed. The harmonic oscillator approximation will eventually break down with increasing off-line excitations and thus increasing $\ell$. 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 (36.12), i.e., the zeros closest to the real energy axis, it is sufficient to take only the $k = 0$ terms into account.

Next, let us have a look at the discrete symmetries discussed in sect. 36.2. Collinear helium has a $C_2$ symmetry as it is invariant under reflection across the $r_1 = r_2$ line corresponding to the electron-electron exchange symmetry. As explained in sects. 19.1.1 and 19.5, 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

$$1/\zeta^{(\ell)} = \prod_a (1 - t_a)^2 \prod_{\tilde{s}} (1 - t_{\tilde{s}}^2) \,. \tag{36.14}$$

Here, the first product is taken over all asymmetric prime cycles, i.e., cycles that are not self-dual under the $C_2$ 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 $t_{\tilde{s}}$ in (36.14) are the weights of fundamental domain cycles. The $C_2$ symmetry now leads to the factorization of (36.14) $1/\zeta = \zeta_+^{-1}\zeta_-^{-1}$, with

$$1/\zeta_+^{(\ell)} = \prod_a (1 - t_a) \prod_{\tilde{s}} (1 - t_{\tilde{s}}) \,,$$

$$1/\zeta_-^{(\ell)} = \prod_a (1 - t_a) \prod_{\tilde{s}} (1 + t_{\tilde{s}}) , \tag{36.15}$$

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. 36.2. An orbit with an odd number of 1's in the itinerary is self-dual under the $C_2$ symmetry and enters the spectral determinant in (36.15) with a negative or a positive sign, depending on the symmetry subspace under consideration.

### 36.5.3 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 $\pm 1$ corresponds to states being symmetric or antisymmetric with respect to the electron-electron exchange. We may now start writing down the binary cycle expansion (18.7) 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 $\overline{0}$ 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 23. We note that the behavior of cycles going far out in the channel $r_1$ or $r_2 \to \infty$ 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 $r_1, r_2 \to \infty$. 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. With this observation in mind, we may write the cycle expansion (....) for a binary alphabet without the $\overline{0}$ cycle as [remark 36.5]

$$\begin{aligned} 1/\zeta^\ell(s) = \quad 1 \quad & - t_1^{(\ell)} - t_{01}^{(\ell)} - [t_{001}^{(\ell)} + t_{011}^{(\ell)} - t_{01}^{(\ell)} t_1^{(\ell)}] \\ & - [t_{0001}^{(\ell)} + t_{0011}^{(\ell)} - t_{001}^{(\ell)} t_1^{(\ell)} + t_{0111}^{(\ell)} - t_{011}^{(\ell)} t_1^{(\ell)}] - \dots . \end{aligned} \tag{36.16}$$

The weights $t_p^{(\ell)}$ are given in (36.12), with contributions of orbits and composite orbits of the same total symbol length collected within square brackets. The cycle expansion depends only on the classical actions, stability indices and winding numbers, given for orbits up to length 6 in table **??**. To get reacquainted with the

| $N$ | $n$ | $j=1$ | $j=4$ | $j=8$ | $j=12$ | $j=16$ | $-E_{\rm qm}$ |
|---|---|---|---|---|---|---|---|
| 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 |

**Table 36.2:** Collinear helium, real part of the symmetric subspace resonances obtained by a cycle expansion (36.16) up to cycle length $j$. The exact quantum energies [3] 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.

cycle expansion formula (36.16), 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

$$E_{m,N} = -\frac{(S_1/2\pi)^2}{[m+\frac{1}{2}+2(N+\frac{1}{2})\sigma_1]^2}, \qquad m, N = 0, 1, 2, \ldots \,, \tag{36.17}$$

with $S_1/2\pi = 1.8290$ for the action and $\sigma_1 = 0.5393$ for the winding number, see table **??**, the 1 cycle in the fundamental domain. This cycle can be described as the *asymmetric stretch* orbit, see figure 36.5. The additional quantum number $N$ in (36.17) corresponds to the principal quantum number defined in sect. 36.5.1. The states described by the quantization condition (36.17) 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 36.6. The simple formula (36.17) gives already a rather good estimate for the ground state of helium! Results obtained from (36.17) are tabulated in table 36.2, 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 (36.16), 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 36.2 for symmetric states obtained from truncations of the cycle expansion of $1/\zeta_+$.

[exercise 36.7]

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 (36.15)

reveals states in the Rydberg series which are to the right of the energetically lowest series in figure 36.6.

## Résumé

We have covered a lot of ground starting with considerations of the classical properties of a three-body Coulomb problem, and ending with the semiclassical 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 needed 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 $\ell$. We thus understood the origin of the different Rydberg series depicted in figure36.6, a fact which is not at all obvious from a numerical solution of the quantum problem.

Having traversed the long road from the classical 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 now see, the periodic orbit theory has still much of interest to offer.

## Commentary

**Remark 36.1** 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. [2]. The general two–body collision regularizing Kustaanheimo–Stiefel transformation [5], a generalization of Levi-Civita's [13] Pauli matrix two–body collision regularization for motion in a plane, is due to Kustaanheimo [12] who realized that the correct higher-dimensional generalization of the "square root removal" trick (6.15), 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 [6] to obtain numerically stable solutions for planetary motions. The basic idea was in place as early as 1931, when H. Hopf [14] used a KS transformation in order to illustrate a Hopf's invariant. The KS transformation for the collinear helium was introduced in ref. [2].

**Remark 36.2** 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 dynamcis, 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. [2].

**Remark 36.3** 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 36.4** Helium quantum numbers. The classification of the helium states in terms of single electron quantum numbers, sketched in sect. 36.5.1, 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 $\Theta$, see figure 36.1, 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. 36.5.1 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 classical 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 classical $\Theta = 0$ dynamics are on the outermost rigth side in each $N$ subspectrum in figure 36.6, and contain the energetically highest states for given $N, n$ quantum numbers, see also remark 36.5. A detailed account of the historical development as well as a modern interpretation of the spectrum can be found in ref. [1].

**Remark 36.5** Beyond the unstable collinear helium subspace. The semiclassical quantization of the chaotic collinear helium subspace is discussed in refs. [7, 8, 9]. Classical and semiclassical considerations beyond what has been discussed in sect. 36.5 follow several other directions, all outside the main of this book.

A classical 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 36.6. Details can be found in refs. [10, 11].

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 symbol sequences 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 $\sum_{n=0}^{\infty} t_{a0^n}$, where $t_p$ are the cycle weights in (36.12), and cycle expansion of indeed yield all Rydberg states up the various ionization thresholds, see ref. [4]. For a comprehensive overview on spectra of two-electron atoms and semiclassical treatments ref. [1].

## Exercises

36.1. **Kustaanheimo–Stiefel transformation.** Check the Kustaanheimo–Stiefel regularization for collinear helium; derive the Hamiltonian (6.18) and the collinear helium equations of motion (6.19).

36.2. **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 $\Theta$, see figure 6.1.

36.3. **Helium trajectories.** Do some trial integrations of the collinear helium equations of motion (6.19). 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-$d$ Poincaré section, exercise 36.4. Most trajectories will run away, do not be surprised - the classical collinear helium is unbound. Try to guess approximately the shortest cycle of figure 36.4.

36.4. **A Poincaré section for collinear Helium.** Construct a Poincaré section of figure 36.3b 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 the figure 36.3a 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 36.5.

36.5. **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 (36.16).

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 length 5 by a 1-dimensional Newton search.

36.6. **Collinear helium cycle stabilities.** Compute the eigenvalues for the cycles you found in exercise 36.5, as described in sect. 36.3. You may either integrate the reduced $2 \times 2$ matrix using equations (36.6) together with the generating function $\mathbf{I}$ given in local coordinates by (36.7) or integrate the full $4 \times 4$ Jacobian matrix, see sect. 22.1. 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 **??**; you should get the actions and Lyapunov exponents right, but topological indices and stability angles we take on faith.

36.7. **Helium eigenenergies.** Compute the lowest eigenenergies of singlet and triplet states of helium by substituting cycle data into the cycle expansion (36.16) for the symmetric and antisymmetric zeta functions (36.15). 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 18.1, can yield informed guesses. Better way would be to find the zeros by Newton method, sect. 18.2.3. How close are you to the cycle expansion and quantum results listed in table 36.2? You can find more quantum data in ref. [3].

## References


[36.1] G. Tanner, J-M. Rost and K. Richter, Rev. Mod. Phys. **72**, 497 (2000).

[36.2] K. Richter, G. Tanner, and D. Wintgen, Phys. Rev. A **48**, 4182 (1993).

[36.3] Bürgers A., Wintgen D. and Rost J. M., J. Phys. B **28**, 3163 (1995).

[36.4] G. Tanner and D. Wintgen Phys. Rev. Lett. **75** 2928 (1995).

[36.5] P. Kustaanheimo and E. Stiefel, *J. Reine Angew. Math.* **218**, 204 (1965).

[36.6] E.L. Steifel and G. Scheifele, *Linear and regular celestial mechanics* (Springer, New York 1971).

[36.7] G.S. Ezra, K. Richter, G. Tanner and D. Wintgen, J. Phys. B **24**, L413 (1991).

[36.8] D. Wintgen, K. Richter and G. Tanner, CHAOS **2**, 19 (1992).

[36.9] R. Blümel and W. P. Reinhardt, *Directions in Chaos Vol 4*, eds. D. H. Feng and J.-M. Yuan (World Scientific, Hongkong), 245 (1992).

[36.10] K. Richter and D. Wintgen, J. Phys. B **24**, L565 (1991).

[36.11] D. Wintgen and K. Richter, Comments At. Mol. Phys. **29**, 261 (1994).

[36.12] P. Kustaanheimo, *Ann. Univ. Turku, Ser. AI.*, **73** (1964).

[36.13] T. Levi-Civita, *Opere mathematische* **2** (1956).

[36.14] H. Hopf, *Math. Ann.* **104** (1931).

# Chapter 37

# Diffraction distraction

(N. Whelan)

DIFFRACTION EFFECTS characteristic to scattering off wedges are incorporated into the periodic orbit theory.

## 37.1 Quantum eavesdropping

As noted in chapter 36, the classical 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 classical 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 chapter34. 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 classical 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.

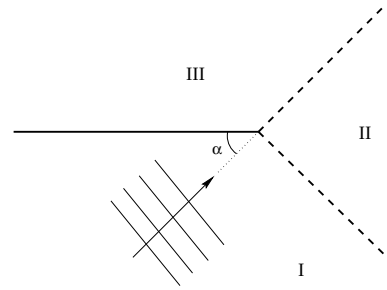



**Figure 37.1:** Scattering of a plane wave off a half line.

We start by solving the problem of the scattering of a plane wave off a half line (see figure 37.1). 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 $\alpha$. What is the total field? This is a problem solved by Sommerfeld 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

$$v(r,\psi) = Ae^{-ikr\cos\psi} \tag{37.1}$$

where $\psi = \phi - \alpha$ and $\phi$ is the angular coordinate. The total field is then given by the method of images as

$$v_{\text{tot}} = v(r,\phi-\alpha) - v(r,\phi+\alpha), \tag{37.2}$$

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

$$v(r,\psi) = A\int_C d\beta\, f(\beta,\psi)e^{-ikr\cos\beta}. \tag{37.3}$$

This is certainly correct if the function $f(\beta,\psi)$ has a pole of residue $1/2\pi i$ at $\beta =$

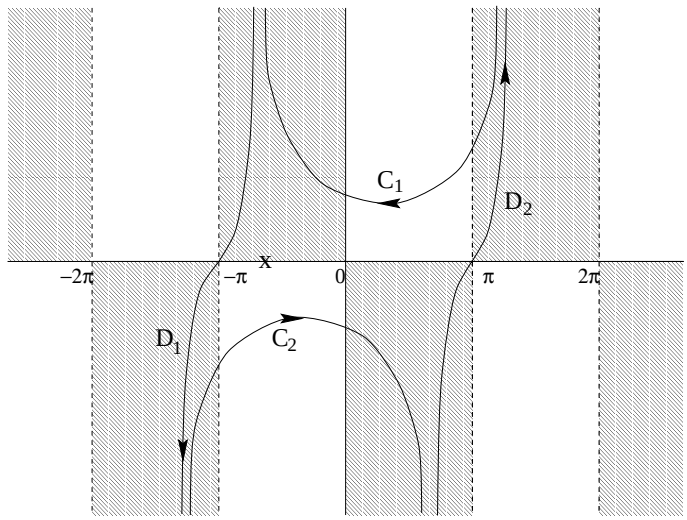



**Figure 37.2:** The contour in the complex $\beta$ plane. The pole is at $\beta = -\psi$ (marked by $\times$ in the figure) and the integrand approaches zero in the shaded regions as the magnitude of the imaginary part of $\beta$ approaches infinity.

$-\psi$ and if the contour $C$ encloses that pole. One choice is

$$f(\beta, \psi) = \frac{1}{2\pi} \frac{e^{i\beta}}{e^{i\beta} - e^{-i\psi}}. \tag{37.4}$$

(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 37.2. This encloses the pole and vanishes as $|\mathrm{Im}\,\beta| \to \infty$ (as denoted by the shading). The sections $D_1$ and $D_2$ are congruent because they are displaced by $2\pi$. 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 $2\pi$ periodicity of the problem. The half line scattering problem follows by analogy.

Whereas for the full line the field is periodic in $2\pi$, for the half line it is periodic in $4\pi$. This can be seen by the fact that the field can be expanded in a series of the form $\{\sin(\phi/2), \sin(\phi), \sin(3\phi/2), \cdots\}$. As above, we extend the space by thinking of it as two sheeted. The physical sheet is as shown in figure 37.1 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 $r \to \infty$ 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

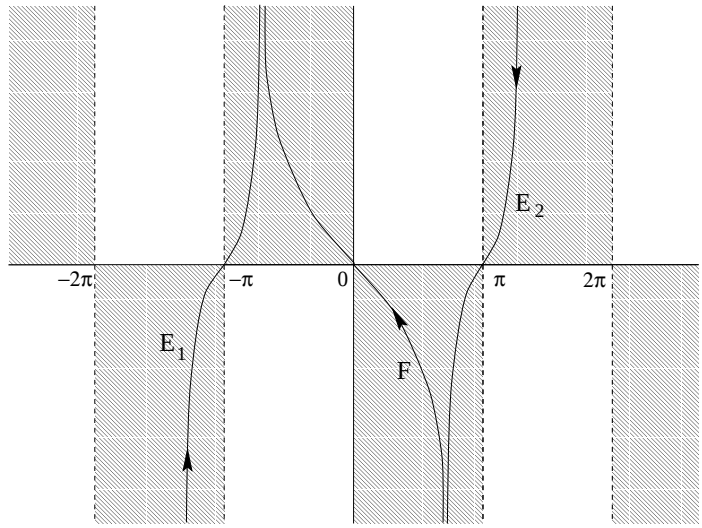



**Figure 37.3:** The contour used to evaluate the diffractive field after the contribution of possible poles has been explicitly evaluated. The curve $F$ is traversed twice in opposite directions and has no net contribution.

as

$$v_{\text{tot}} = u(r, \phi - \alpha) - u(r, \phi + \alpha). \tag{37.5}$$

where $u(r, \psi)$ is a $4\pi$-periodic function to be determined. The second term is 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 (37.3) with the same contour $C_1 + C_2$ but with the $4\pi$ periodicity accounted for by replacing equation (37.4) with

$$f(\beta, \psi) = \frac{1}{4\pi} \frac{e^{i\beta/2}}{e^{i\beta/2} - e^{-i\psi/2}}. \tag{37.6}$$

(We divide by $4\pi$ rather than $2\pi$ so that the residue is properly normalized.) The integral (37.3) 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 (37.4). 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 $\phi$ 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 37.2.

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 37.2 but with a reversed sense. Since the integrand is no longer $2\pi$ periodic, the contributions from these curves no longer cancel. We evaluate both stationary phase integrals to obtain

$$u(r,\psi) \approx -A\frac{e^{i\pi/4}}{\sqrt{8\pi}}\sec(\psi/2)\frac{e^{ikr}}{\sqrt{kr}} \tag{37.7}$$

so that the total diffracted field is

$$v_{\text{diff}} = -A\frac{e^{i\pi/4}}{\sqrt{8\pi}}\left(\sec\left(\frac{\phi-\alpha}{2}\right) - \sec\left(\frac{\phi+\alpha}{2}\right)\right)\frac{e^{ikr}}{\sqrt{kr}}. \tag{37.8}$$

Note that this expression breaks down when $\phi \pm \alpha = \pi$. These angles correspond to the borders among the three regions of figure 37.1 and must be handled more carefully - we can not do a stationary phase integral in the vicinity of a pole. However, the integral representation (37.3) and (37.6) is uniformly valid. [exercise 37.1]

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

$$G_{\text{f}}(x_2, x_1, k) = -\frac{i}{4}H_0^{(+)}(kd) \approx -\frac{1}{\sqrt{8\pi kd}}\exp\{i(kd+\pi/4)\}, \tag{37.9}$$

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 (37.8), 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

$$G_{\text{diff}}(x, x', k) = G_{\text{f}}(x, x_V, k)d(\theta,\theta')G_{\text{f}}(x_V, x', k), \tag{37.10}$$

where, by comparison with equations (37.8) and (37.9), we have

$$d(\theta,\theta') = \sec\left(\frac{\theta-\theta'}{2}\right) - \sec\left(\frac{\theta+\theta'}{2}\right). \tag{37.11}$$

Here $\theta'$ is the angle to the source as measured from the vertex and $\theta$ is the angle to the receiver. They were denoted as $\alpha$ and $\phi$ 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 $G_{\text{diff}}$ is that it is a simple multiplicative combination of other semiclassical Green's functions.

[exercise 37.2]

We now recover our classical perspective by realizing that we can still think of classical trajectories. In calculating the quantum Green's function, we sum over the contributions of various paths. These include the classical trajectories which connect the points and also paths which connect the points via the vertex. These have different weights as given by equations (37.9) and (37.10) but the concept of summing over classical 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 [15]. It can be written as

$$G(x, x', k) = g(r, r', k, \theta' - \theta) - g(r, r', k, \theta' + \theta) \tag{37.12}$$

where $(r, \theta)$ and $(r', \theta')$ 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

$$g(r, r', k, \psi) = \frac{i}{8\pi\nu} \int_{C_1+C_2} d\beta \frac{H_0^+(k\sqrt{r^2 + r'^2 - 2rr'\cos\beta})}{1 - \exp\left(i\frac{\beta+\psi}{\nu}\right)} \tag{37.13}$$

where $\nu = \gamma/\pi$ and $\gamma$ 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 37.2.

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

$$d(\theta, \theta') = -\frac{4\sin\frac{\pi}{\nu}}{\nu} \frac{\sin\frac{\theta}{\nu}\sin\frac{\theta'}{\nu}}{\left(\cos\frac{\pi}{\nu} - \cos\frac{\theta+\theta'}{\nu}\right)\left(\cos\frac{\pi}{\nu} - \cos\frac{\theta-\theta'}{\nu}\right)}, \tag{37.14}$$

which reduces to (37.11) 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

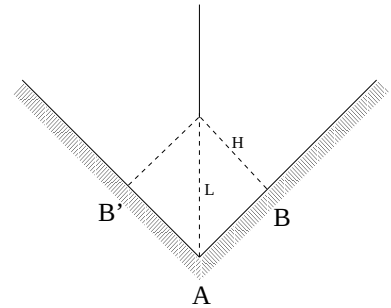



**Figure 37.4:** The billiard considered here. The dynamics consists of free motion followed by specular reflections off the faces. The top vertex induces diffraction while the bottom one is a right angle and induces two specular geometric reflections.

contribution to any order. Classically 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 classical 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 $\pi/n$. For these geometries the saddle point analysis leading to (37.14) 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.

## 37.2 An application

Although we introduced diffraction as a correction to the purely classical 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 37.4 The classical 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 34 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 classical trajectories. In general, $g(k)$ is given by the sum over all diffractive and geometric orbits. The contribution of the simple diffractive orbit labeled $\gamma$ shown in figure 37.5 to $g(k)$ is determined as follows.

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 (37.9) and (37.10). 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 $2ikL+iky^2/(L^2-x^2)$.

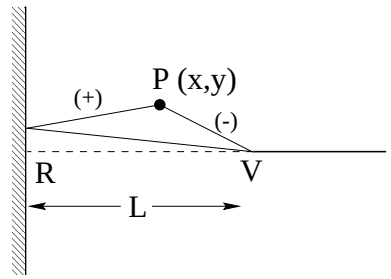



**Figure 37.5:** The dashed line shows a simple periodic diffractive orbit $\gamma$. Between the vertex $V$ and a point $P$ close to the orbit there are two geometric legs labeled $\pm$. The origin of the coordinate system is chosen to be at $R$.

The trace integral involves integrating over all points $P$ and is

$$g_\gamma(k) \approx -2d_\gamma \frac{e^{i(2kL+\pi/2)}}{8\pi k} \int_0^L \frac{dx}{\sqrt{L^2-x^2}} \int_{-\infty}^{\infty} dy e^{\left(iky^2 \frac{L}{L^2-x^2}\right)}. \tag{37.15}$$

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 $d_\gamma$ is the one corresponding to the diffractive periodic orbit. To evaluate the $y$ integral, we use the identity

$$\int_{-\infty}^{\infty} d\xi e^{ia\xi^2} = e^{i\pi/4}\sqrt{\frac{\pi}{a}}, \tag{37.16}$$

and thus obtain a factor which precisely cancels the $x$ dependence in the $x$ integral. This leads to the rather simple result

$$g_\gamma \approx -\frac{il_\gamma}{2k}\left\{\frac{d_\gamma}{\sqrt{8\pi k l_\gamma}}\right\} e^{i(kl_\gamma+\pi/4)} \tag{37.17}$$

where $l_\gamma = 2L$ is the length of the periodic diffractive orbit. A more sophisticated analysis of the trace integral has been done [6] using the integral representation (37.13). It is valid in the vicinity of an optical boundary and also for wedges with opening angles close to $\pi/n$.

Consider a periodic diffractive orbit with $n_\gamma$ reflections off straight hard walls and $\mu_\gamma$ 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 $l_\gamma$ is the length of the corresponding prime orbit. For such an orbit, (37.17) generalizes to

$$g_\gamma(k) = -\frac{il_\gamma}{2k}\left\{\prod_{j=1}^{\mu_\gamma} \frac{d_{\gamma,j}}{\sqrt{8\pi k l_{\gamma,j}}}\right\} \exp\{i(kL_\gamma + n_\gamma\pi - 3\mu_\gamma\pi/4)\}. \tag{37.18}$$

[exercise 37.3]

Each diffraction introduces a factor of $1/\sqrt{k}$ and multi-diffractive orbits are thereby suppressed.

If the orbit $\gamma$ is prime then $L_\gamma = l_\gamma$. If $\gamma$ is the $r$'th repeat of a prime orbit $\beta$ we have $L_\gamma = rl_\beta$, $n_\gamma = rp_\beta$ and $\mu_\gamma = r\sigma_\beta$, where $l_\beta$, $p_\beta$ and $\sigma_\beta$ all refer to the prime orbit. We can then write

$$g_\gamma = g_{\beta,r} = -\frac{il_\beta}{2k} t_\beta^r \tag{37.19}$$

where

$$t_\beta = \left\{ \prod_{j=1}^{\sigma_\beta} \frac{d_{\beta,j}}{\sqrt{8\pi k l_{\beta,j}}} \right\} \exp\{i(kl_\beta + p_\beta\pi - 3\sigma_\beta\pi/4)\}. \tag{37.20}$$

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

$$g_{\text{diff}}(k) = \sum_\beta \sum_{r=1}^{\infty} g_{\beta,r} = -\frac{i}{2k} \sum_\beta l_\beta \frac{t_\beta}{1-t_\beta}. \tag{37.21}$$

We cast this as a logarithmic derivative (17.7) by noting that $\frac{dt_\beta}{dk} = il_\beta t_\beta - \sigma_\beta t_\beta/2k$ and recognizing that the first term dominates in the semiclassical limit. It follows that

$$g_{\text{diff}}(k) \approx \frac{1}{2k}\frac{d}{dk}\left\{ \ln \prod_\beta (1-t_\beta) \right\}. \tag{37.22}$$

In the case that there are only diffractive periodic orbits - as in the geometry of figure 37.4 - the poles of $g(k)$ are the zeros of a dynamical zeta function

$$1/\zeta(k) = \prod_\beta (1-t_\beta). \tag{37.23}$$

For geometric orbits, this function would be evaluated with a cycle expansion as discussed in chapter 18. However, here we can use the multiplicative nature of the weights $t_\beta$ to find a closed form representation of the function using a Markov graph, as in sect. 10.4.1. This multiplicative property of the weights follows from the fact that the diffractive Green's function (37.10) 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 sects. 19.5 and 19.1.1) and we determine $1/\zeta_+$ and $1/\zeta_-$ separately using the ideas of symmetry decomposition of chapter 19.

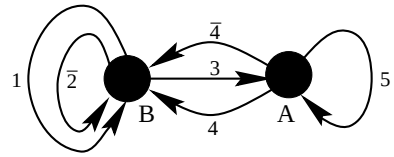



**Figure 37.6:** The two-node Markov graph with all the diffractive processes connecting the nodes.

In the Markov graph shown in figure 37.6, we enumerate all processes. We start by identifying the fundamental domain as just the right half of figure 37.4. 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 37.6. 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. 19.5 we find

$$\begin{aligned} 1/\zeta_+ &= 1 - t_1 - t_{\bar{2}} - t_5 - t_3 t_4 - t_3 t_{\bar{4}} + t_5 t_1 + t_5 t_{\bar{2}}\,, \\ &= 1 - (t_1 + t_{\bar{2}} + t_5) - 2 t_3 t_4 + t_5 (t_1 + t_{\bar{2}}) \end{aligned} \tag{37.24}$$

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. 19.3.1) and only affects the even resonances - the terms involving $t_5$ are absent from $1/\zeta_-$. The result is

$$\begin{aligned} 1/\zeta_- &= 1 - t_1 + t_{\bar{2}} - t_3 t_4 + t_3 t_{\bar{4}}\,, \\ &= 1 - (t_1 - t_{\bar{2}}). \end{aligned} \tag{37.25}$$

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 37.4]

It now just remains to fix the weights. We use equation (37.20) but note that each weight involves just one diffraction constant. It is then convenient to define the quantities

$$u_A^2 = \frac{\exp\{i(2kL + 2\pi)\}}{\sqrt{16\pi k L}} \qquad u_B^2 = \frac{\exp\{i(2kH + \pi)\}}{\sqrt{16\pi k H}}. \tag{37.26}$$

The lengths $L$ and $H = L/\sqrt{2}$ are defined in figure 37.4; we set $L = 1$ throughout. Bouncing inside the right angle at $A$ corresponds to two specular reflections so that

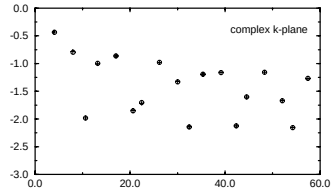



**Figure 37.7:** The even resonances of the wedge scatterer of figure 37.4 plotted in the complex $k$–plane, with $L = 1$. The exact resonances are represented as circles and their semiclassical approximations as crosses.

$p = 2$. We therefore explicitly include the factor $\exp(i2\pi)$ in (37.26) 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 (37.11) 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:

$$\begin{aligned} t_1 = sd_{BB}u_B^2 \quad t_{\bar{2}} &= sd_{B'B}u_B^2 \quad t_3 = t_4 = t_{\bar{4}} = sd_{AB}u_Au_B \\ t_5 &= sd_{AA}u_A^2. \end{aligned} \tag{37.27}$$

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 $t_1 = t_{\bar{2}}$. From (37.25) this means that there are no odd resonances because 1 can never equal 0. For the even resonances equation (37.24) is an implicit equation for $k$ which has zeros shown in figure 37.7.

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 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 37.5]

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 [7] 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 37.4 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 classical orbits. They are generated by singularities of the scattering potential. In these singular points the classical dynamics has no unique definition, and the classical orbits hitting the singularities can be continued in many different directions. While the classical 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 classically 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 37.1** Classical 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. [1, 2, 3, 7]
- a vertex such that trajectories to one side bounce differently from those to the other side, refs. [2, 4, 5, 8, 9].
- a point scatterer [10, 11] or a magnetic flux line [12, 13] such that we do not know how to continue classical mechanics through the discontinuities.

**Remark 37.2** 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 (37.8). 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 37.3** Generalizations Consider the effect of replacing our half line by a wedge of angle $\gamma_1$ and the right angle by an arbitrary angle $\gamma_2$. If $\gamma_2 > \gamma_1$ and $\gamma_2 \geq \pi/2$ this is an open problem whose solution is given by equations (37.24) and (37.25) (there will then be odd resonances) but with modified weights reflecting the changed geometry [8]. (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 [21, 22]. 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 (37.7) is invalid. Even the more uniform derivation of [6] 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 37.4** Diffractive Green's functions. The result (37.17) is proportional to the length of the orbit times the semiclassical Green's function (37.9) to go from the vertex back to itself along the classical path. The multi-diffractive formula (37.18) 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 classical 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 [2] and a derivation can be found in reference [9]. A similar structure also exists for creeping [2].

**Remark 37.5** 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 [23]. 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 36. 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 classical 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 [24].

## Exercises

37.1. **Stationary phase integral.** Evaluate the two stationary phase integrals corresponding to contours $E_1$ and $E_2$ of figure 37.3 and thereby verify (37.7).

(N. Whelan)

37.2. **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

$$d = \frac{2\pi}{\log\left(\frac{2}{ka}\right) - \gamma_e + i\frac{\pi}{2}} \tag{37.28}$$

where $\gamma_e = 0.577\cdots$ is Euler's constant. Note that in this limit $d$ depends weakly on $k$ but not on the scattering angle.

(N. Whelan)

37.3. **Several diffractive legs.** Derive equation (37.18). 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)

37.4. **Unsymmetrized dynamical zeta function.** Assume you know nothing about symmetry decomposition. Construct the three node Markov diagram for figure 37.1 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 the even and odd dynamical zeta functions. Why is there no term $t_{\bar{2}}$ in the full dynamical zeta function?

(N. Whelan)

37.5. **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 37.2 to construct the desymmetrized dynamical zeta functions, as in sect. 19.6. 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 10. By assuming that the real part of $k$ is much greater than the imaginary part show that the positions of the resonances are $k_nR = \alpha_n - i\beta_n$ where $\alpha_n = 2\pi n + 3\pi/4$, $\beta_n = \log\left(\sqrt{2\pi\alpha_n}/d\right)$ and $n$ is a non-negative integer. (See also reference [11].)

(N. Whelan)

## References


[37.1] A. Wirzba, CHAOS **2**, 77 (1992);

[37.2] G. Vattay, A. Wirzba and P. E. Rosenqvist, Phys. Rev. Lett. **73**, 2304 (1994); G. Vattay, A. Wirzba and P. E. Rosenqvist in *Proceedings of the International Conference on Dynamical Systems and Chaos: vol. 2*, edited by Y.Aizawa, S.Saito and K.Shiraiwa (World Scientific, Singapore, 1994).

[37.3] H. Primack, H. Schanz, U. Smilansky and I. Ussishkin, Phys. Rev. Lett. **76**, 1615 (1996).

[37.4] N. D. Whelan, Phys. Rev. E **51**, 3778 (1995).

[37.5] N. Pavloff and C. Schmit, Phys. Rev. Lett. **75**, 61 (1995).

[37.6] M. Sieber, N. Pavloff, C. Schmit, Phys. Rev. E **55**, 2279 (1997).

[37.7] H. Primack et. al., J. Phys. A **30**, 6693 (1997).

[37.8] N. D. Whelan, Phys. Rev. Lett. **76**, 2605 (1996).

[37.9] H. Bruus and N. D. Whelan, Nonlinearity, **9**, 1 (1996).

[37.10] P. Seba, Phys. Rev. Lett. **64**, 1855 (1990).

[37.11] P. E. Rosenqvist, N. D. Whelan and A. Wirzba, J. Phys. A **29**, 5441 (1996).

[37.12] M. Brack et. al., Chaos **5**, 317 (1995).

[37.13] S. M. Reimann et. al., Phys. Rev. A **53**, 39 (1996).

[37.14] A. Sommerfeld, Mathem. Ann. **47**, 317 (1896); *Optics* (Academic Press, New York 1954).

[37.15] H. S. Carslaw, Proc. London Math. Soc. (Ser. 1) **30**, 121 (1989); H. S. Carslaw, Proc. London Math. Soc. (Ser. 2) **18**, 291 (1920).

[37.16] J. B. Keller, J. Appl. Phys. **28**, 426 (1957).

[37.17] A. Voros, J. Phys. A **21**. 685 (1988).

[37.18] see for example, D. Ruelle, *Statistical Mechanics, Thermodynamic Formalism* (Addison-Wesley, Reading MA, 1978).

[37.19] see for example, P. Grassberger, Z. Naturforsch. **43a**, 671 (1988).

[37.20] P. Cvitanović and B. Eckhardt, Nonlinearity **6**, 277 (1993).

[37.21] P. Exner, P. Seba and P. Stovicek, Czech J. Phys **B39**, 1181 (1989).

[37.22] Hua Wu and D. W. L. Sprung, J. Appl. Phys. **72**, 151 (1992).

[37.23] P. A. Dando, T. S. Monteiro, D. Delande and K. T. Taylor, Phys. Rev. Lett. **74**, 1099 (1995). P. A. Dando, T. S. Monteiro and S. M. Owen, preprint (1997).

[37.24] D. Delande et. al., J. Phys. B **27**, 2771 (1994); G. Raithel et. al., J. Phys. B **27**, 2849 (1994); M. Courtney et. al., Phys. Rev. Lett., **73**, 1340 (1994).

# 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 méchanique 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 lengths 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 cycle points of $f$ into nearby cycle 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 fundamental matrixs $df^n(x)/dx$ 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, as we shall see in appendix I.1.2, and for smooth flows the convergence of cycle expansions can even be super-exponential.

The cycle expansions such as (18.7) outperform the pedestrian methods such as extrapolations from the finite cover sums (20.2) 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 (20.2) is not. Cycles are of finite period but infinite duration, so the cycle eigenvalues are already evaluated in the $n \to \infty$ limit, but for the sum (20.2) the limit has to be estimated by numerical extrapolations. And, crucially, the higher terms in the cycle expansion (18.7) 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 curtain 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. **Non-hyperbolicity** 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 classically 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.

# Index


action, 232
admissible
    trajectories, number of, 192
alphabet, 137
arc, 139
area preserving
    Hénon map, 96
attractor, 70
    basin, 70
    Hénon, 45
    strange, 70, 233
average
    space, 216, 232
    time, 232
averaging, 190

baker's map, 107
basin of attraction, 70
bi-infinite itinerary, 139
bifurcation
    generic, 107
    saddle-node, 43
billiard, 107
    map, 107
    stability, 107
    stadium, 107, 108
Birkhoff
    coordinates, 108
block
    finite sequence, 139
block, pruning, 139
Boltzmann, L, 167
Bowen, R., 216
brain, rat, 19, 229
butterfly effect, 43

$C_{3v}$ symmetry, 156
Cartwright, M.L., 117
ceiling function, 249
change
    of coordinates, 83
chaos, 48, 52
    caveats, 76
    deterministic, 205
    diagnostics, 105
    quantum, 211
    skeleton of, 104
characteristic
    exponent, 55
    value, 55
chicken heart palpitations, 52
coarse-graining, 216
combinatorics
    teaching, 138
complete
    $N$-ary dynamics, 137
completeness
    relation, 56
complex eigenvalues, 56
confession
    C.N. Yang, 216
    St. Augustine, 216
conjugacy, 83
    smooth, 83, 86
    topological, 138
contracting
    Floquet multipliers, 74
    flow, 57, 70
coordinate
    change, 83
    transformations, 86
Copenhagen School, xi
covering
    symbolic dynamics, 139
critical
    point, 74, 138
    value, 138
cumulant
    expansion, 193
curvature
    expansion, 186
cycle, 105
    fundamental, 192
    limit, 233
    Lyapunov exponent, 74
    marginal stability, 117
    point, 105, 139, 148
        count, 193
    prime, 139, 176, 195, 248
        3-disk, 176
    pruning, 193
    Rössler
        flow, 176
    stability, 74–75

stable, 74
superstable, 74

degree of freedom, 78, 96
density
evolution, 163
phase space, 217
determinant
for flows, 259
graph, 194
Hadamard, 259
spectral, 157, 192, 259
deterministic dynamics, 35, 49, 217
differential equations
ordinary
almost, 108
diffusion
constant, 233
Diffusion Limited Aggregates, 226
dike map, 138
dimension
intrisic, 78
Dirac delta derivatives, 218
Dirac delta function, 218, 233, 248
Duffing oscillator, 91, 96, 105
dynamical
system, 28, 44
deterministic, 35
gradient, 108
smooth, 44
systems
equivalent, 85
dynamical system
smooth, 137, 186, 193
dynamics
deterministic, 49
irreversible, 84
reversible, 86
spatiotemporal, 209
stochastic, 52
symbolic, 95, 139
symmetry, 114, 116
topological, 139

edge, 139
eigenvalues
complex, 56
English
plain, 139
entropy
Kolmogorov, 107, 194
topological, 58, 192–194
equilibrium
Lorenz flow, 95
point, 91
Rössler flow, 56, 109
equivalence
of dynamical systems, 85
equivariance, 114
equivariant, 116
ergodic
theorem
multiplicative, 235
escape
rate, 113
Eulerian
coordinates, 54
evolution
group, 108
kernel
probabilistic, 217
operator, 147
semigroup, 233
expanding
Floquet multipliers, 74
exponent
Floquet, 74
exponential proliferation, 148

factor group, 115
first return function, 41
fixed point
maps, 45
Floquet
exponents, 74
multiplier, 55, 74
flow, 26–101
contracting, 57, 70
deterministic, 217
elliptic, 74
Hamiltonian, 96
hyperbolic, 74
incompressible, 57
linear, 55, 58
linearized, 54
nonhyperbolic, 74
spectral determinant, 259
stability, 56
stochastic, 217
stretch & fold, 138
form
normal, 84
fractal
aggregates, 86
geometry of nature, 86
probabilistic, 86
frequency analysis, 105
functional, 216
composition, 82
Lyapunov, 70
fundamental
cycle, 192

matrix, 54, 124

Gatto Nero
professor, 138
Gauss map, 218
generating function, 248
generating partition, 139
Gilmore, R., 117
gradient
system, 108
grammar
symbolic dynamics, 139
graph, 139
Markov, 139
group
dynamical, 85
evolution, 108
finite, 114
order of, 114

Hadamard determinant, 259
Hamiltonian
dynamics, 96–97
flow
stability, 96
Hénon map, 96
helium
collinear, 45, 96, 109
Hénon map, 43, 157
attractor, 45
fixed points, 45
Hamiltonian, 96
Hénon, M., 43
Hénon-Heiles
symbolic dynamics, 117
heroes
unsung, xi, xxiv
Hilbert-Weyl theorem, 114
horseshoe
complete, 157
hyperbolic
flow, 74
non-, 172
hyperbolicity assumption, 128, 248

incommesurate, 60
incompressible flow, 57
indecomposability
metric, 137
initial
point $x_0$, 50, 54, 123
state $x_0$, 50, 123
integrable system, 83, 96
integrated observable, 232, 233, 248
integration
Runge-Kutta, 109
intermittency, 107
invariance
of flows, 74
invariant
measure, 216
metric, 74
topological, 74
invariant measure
Gauss map, 218
invariant subgroup, 115
inverse iteration, 177
irreducible
segment, 115
irreversibility, 84, 165
isotropy, 117
isotropy subgroup, 114
iteration, 48
inverse, 177
map, 43
itinerary, 42, 98, 137
bi-infinite, 137, 139
future, 139
past, 139

Jacobi, C.G.J., 58
Jacobian, 57

kneading
determinant, 159
sequence, 138
theory, 138
value, 138, 140
Kolmogorov entropy, 107, 194
Kustaanheimo-Stiefel transformation, 84

Lagrangian
coordinates, 54
Laplace
transform, 193, 248, 249
transform, discrete, 248
Laplace, Pierre-Simon de, 42
least action principle, 176
Leibniz, Gottfried Wilhelm, 42
Letellier, C., 117
level set, 96
Lie
algebra, 117
lifetime, 118
limit
cycle, 233
linear
flow, 55, 58
stability, 54, 74
linearized
flow, 54
link, 139
local
stability, 54, 74

Lorenz flow, 42, 56, 57, 93, 115, 118, 138
polar coordinates, 118
proto-Lorenz, 118
Lorenz, E.N., 43, 117
Lozi map, 43
Lyapunov exponent
cycle, 74
numerical, 235
Lyapunov functional, 70
Lyapunov time, 85

$\mathcal{M}$ state space volume, 233
manifold
stable, 156
map, 43, 48
expanding, 137
fixed point, 45
Hénon, 43
Hamiltonian, 96
Hamiltonian
Hénon, 96
iteration, 43
Lozi, 43
once-folding, 157
return, 41, 127, 157
sawtooth, 115
stability, 57
unimodal, 138
marginal
stability, 74, 117
cycle, 117
Markov
graph, 139
partition
finite, 137
not unique, 156
Maupertuis, P.L.M. de, 176
measure
invariant, 216
natural, 43
smooth, 232
mechanics
statistical, 159
metric
indecomposability, 137
invariant, 74
Mira, C., 43
Misiurewicz, M., 43
mixing, 55, 62, 217
Moebius inversion, 193
multiplicative ergodic theorem, 235
multiplier
Floquet, 55
multipoint shooting method, 177

natural measure, 43, 178
nature
geometry of, 86
neighborhood, 54, 75
Nero, G., 138
Newton method, 177
flows, 177
Newtonian dynamics, 96
node, 139
non-wandering set, 68
nonhyperbolic
flow, 74
normal
divisor, 115
form, 84

observable, 216, 232
integrated, 232, 233, 248
open systems, 113, 233
orbit, 43, 50, 57, 114
inadmissible, 138
periodic, 58, 139
ordering
spatial, 138, 157
ordinary differential equations
almost, 108
orthogonality
relation, 56
Oseledec multiplicative ergodic theorem, 235

palpitations, chicken heart, 52
partition, 137, 139
generating, 139
infinite, 140, 193
Markov, 137
partition function, 235
past topological coordinate, 157
periodic
orbit, 58, 139
condition, 176, 178
extraction, 176–178
inverse iteration, 177
multipoint shooting, 177
Newton method, 177
relative, 117
Perron-Frobenius
matrix, 192
phase space, 35
density, 217
pinball
simulator, 108
plain English, 139
Poincaré invariants, 97
Poincaré return map, 41
cycle, 75
stability, 58

Poincaré section, 41–42, 157
3-disk, 107
hyperplane, 41
Poincaré, H., 19, 67, 119
point
non-wandering, 68
periodic, 105, 139
wandering, 64
Poisson
bracket, 217
Pomeau, Y., 43
potential
problems, 108
pressure, 235
thermodynamic, 235
prime cycle, 139, 176, 195, 248
3-disk, 195
count, 193
ternary, 156
profile
spatial, 49
pruning
block, 139
individual cycles, 193
primary interval, 139
rules, 137

quasiperiodicity, 60
quotient group, 115

rectification
flows, 83
maps, 84
recurrence, 67, 137
relative
periodic orbit, 117
renormalization, 107
repeller, 113
representative point, 35
residue, 96
return map, 41, 127, 157
reversible
dynamics, 86
Rössler
cycles, 176
equilibria, 56, 109
flow, 42, 45, 105, 109, 137, 233
Roux
Henri, 57, 196
Ruelle, D., 216
Runge-Kutta integration, 109

saddle-node bifurcation, 43
sawtooth map, 115
section
Poincaré, 41, 107
self-similar, 151
semigroup
dynamical, 85
evolution, 233
sensitivity to initial conditions, 43, 233, 237
shadowing, 134, 193
shift, 139
full, 139
sub-, 139
Sinai, Ya., 216
singular value decomposition, 55
singular values, 55
skeleton of chaos, 104
Smale
wild idea, 259
Smale, S., 87, 156, 159, 194, 216
smooth, 116
conjugacy, 83, 86
dynamics, 44, 137, 186, 193
measure, 232
potential, 107
space
average, 216
averaging, 232
spatial
profile, 49
spatiotemporal dynamics, 209
spectral
determinant, 157, 192, 259
for flows, 259
spectral decomposition, 56
specular reflection, 107
St. Augustine, 216
stability, 54–58
billiards, 107
elliptic, 248
exact, 75
flow, 56
Hamiltonian flows, 96
linear, 54, 74
maps, 57
marginal, 117
Poincaré map cycle, 75
Poincaré return map, 58
structural, 157, 193
stable
cycle, 74
manifold, 156
stadium billiard, 107, 108
state, 139
set, 137
state space, 35, 41
discretization, 235
volume $\mathcal{M}$, 233
stationary
state, 216

stationary phase, 217
statistical mechanics, 159
stochastic
    dynamics, 52, 217
Stokes theorem, 97
stosszahlansatz, 167
strange
    attractor, 70
strange attractor, 233
stretch & fold, 138
structural stability, 157, 193
subgroup
    isotropy, 114
subshift, 139
    finite type, 139, 157
superstable cycle, 74
symbol square, 157
symbolic
    dynamics
        at a bifurcation, 107
        complete $N$-ary, 137
        covering, 139
symbolic dynamics, 95, 137–139
    3-disk, 118, 137, 237
    coding, 139
    complete, 157
    grammar, 139
    Hénon-Heiles, 117
symmetry, 114–117
    $C_{3v}$, 156
    3-disk, 116, 156
    discrete, 156
    dynamical system, 114, 116
symplectic
    Hénon map, 96
systems
    open, 233

teaching
    combinatorics, 138
template, 176
tent map, 86
ternary
    prime cycles, 156
thermodynamical
    pressure, 235
3-body problem, 83
3-disk
    cycle
        analytically, 118
        count, 117
    geometry, 107
    hyperbolicity, 248
    pinball, 40, 107
    prime cycles, 130, 176, 195
    simulator, 108
    state space, 42
    symbolic dynamics, 118, 137, 237
    symmetry, 116, 156
    transition matrix, 137
time
    arrow of, 164
    average, 232
    ordered integration, 58
    turnover, 56
topological
    conjugacy, 138
    dynamics, 139
    entropy, 58, 192, 193
    future coordinate, 138
    invariant, 74
    parameter, 139
torus, 60
trace
    local, 192
trajectory, 50
    discrete, 43
transfer
    spectrum, 259
transformation
    coordinate, 86
transient, 65, 137
transition matrix, 137, 192
    3-disk, 137
transversality
    condition, 41
turbulence, 70, 80
turnover time, 56

Ulam
    map, tent, 218
Ulam map, 86
unimodal
    kneading value, 140
    well ordered symbols, 140
unimodal map, 138
unstable
    cycle, 74
    manifold, 156
unsung
    heroes, xi, xxiv

vector
    field, 88
vector fields
    singularities, 83
vertex, 139
visitation frequency, 216

wandering point, 64
weight
    multiplicative, 191
well ordered symbols

unimodal, 140
winding number, 96

Yang, C.N., 216
Young, L.-S., 43

# Chaos: Classical and Quantum

# Volume III: Material available on ChaosBook.org

**ChaosBook.org version12, Jun 22 2008** printed June 26, 2008
ChaosBook.org
comments to: predrag [snail] nbi.dk

# Appendix A

# 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.

—M.V. Berry

(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.

### A.0.1 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 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 different times. The first sign on the impossibility of integrating the three-body problem came from a result of Burns that showed that there were no conserved quantities that were polynomial in the momenta and positions. Burns' result did not preclude the possibility of more complicated conserved quantities. This problem was settled by Poincaré and Sundman in two very different ways.

In an attempt to promote the journal *Acta Mathematica*, Mittag-Leffler got the permission of the 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 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, the 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 the 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 Weirstrass in the 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 Weirstrass's expectations, and the series solution did not "considerably extend our understanding of the solar system.' The work that followed from Poincaré did.

### A.0.2 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. To evaluate the heat capacity of even a simple system, Boltzmann had to make a great simplifying assumption of ergodicity: that the dynamical system would 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 twentieth century it has only been shown true for a few systems and wrong for quite a few 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 when he noticed that it was possible to reformulate it using the recently developed methods of Hilbert spaces. This was an important step that showed that it was possible to take a finite-dimensional nonlinear problem and reformulate it as a infinite-dimensional linear problem. This does not make the problem easier, but it does allow one to use a different set of mathematical tools on 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. He proved that if the mechanical system was ergodic, then the computed averages would make sense. Soon afterwards Birkhoff published a much stronger version of the theorem.

### A.0.3 Nonlinear oscillators

The third problem that was very influential in the development of the theory of chaotic dynamical systems was the work on the 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 a clever use of negative friction he created two basic models for the 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, a periodic motion a system goes to regardless of the initial conditions.

## A.1 Chaos grows up

(R. Mainieri)

The theorems of von Neumann and Birkhoff on the ergodic hypothesis were published in 1912 and 1913. This line of enquiry 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 lead 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 that developed from the ergodic hypothesis was in trying to find mechanical systems that are ergodic. An ergodic system could not have stable orbits, as these would break ergodicity. So in 1898 Hadamard 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 temptation for Mary Lucy Cartwright and J.E. Littlewood [15], 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 11]

In Russia, Lyapunov paralleled the methods of Poincaré and initiated the strong Russian dynamical systems school. 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.

## A.2 Chaos with us

(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 collected his results and their development in the 1967 review article on dynamical systems, entitled "Differentiable dynamical systems." 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 emphasis of the paper is on the global properties of the dynamical system, on how to understand the topology of the orbits. Smale's account takes you from a local differential equation (in the form of vector fields) to the global topological description in terms of horseshoes. [chapter 11]

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, where he discusses the entropy of the shift transformation. Kolmogorov went far beyond and suggested a definition of the metric entropy of an area preserving transformation in order to classify Bernoulli shifts. The suggestion was taken by his student Sinai and the results published in 1959. In 1960 Rohlin connected these results to measure-theoretical notions of entropy. The next step was published in 1965 by Adler and Palis, and also Adler, Konheim, McAndrew; 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. It was in the context of studying the entropy associated to a dynamical system that Sinai 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 every one. 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. 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.

A lecture of Smale and the results of Metropolis, Stein, and Stein inspired Feigenbaum to study simple maps. This lead 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 lead 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 use of presentation functions as the dynamics of scaling functions.

The work in more than one dimension progressed very slowly and is still far from completed. 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." Thurston's techniques exposed in that lecture have not been applied in physics, but much of the classification that Thurston developed can be obtained from the notion of a 'pruning front' developed independently by Cvitanović.

Once one develops an understanding for 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 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 will help get a sense of perspective on the field. It is not a fad and it will not die anytime soon.

## A.3 Periodic orbit theory

> Pure mathematics is a branch of applied mathematics.
>
> — Joe Keller, after being asked to define applied mathematics

The history of the periodic orbit theory is rich and curious, and the recent advances are to equal degree inspired by a century of separate development of three disparate subjects; 1. *classical chaotic dynamics*, initiated by Poincaré and put on its modern footing by Smale [23], Ruelle [28], and many others; 2. *quantum theory* initiated by Bohr, with the modern 'chaotic' formulation by Gutzwiller [12, 17]; and 3. analytic *number theory* initiated by Riemann and formulated as a spectral problem by Selberg [20, 3]. Following totally different lines of reasoning and driven by very different motivations, the three separate roads all arrive at formally nearly identical *trace formulas*, *zeta functions* and *spectral determinants*.

That these topics should be related is far from obvious. Connection between dynamics and number theory arises from Selberg's observation that description of geodesic motion and wave mechanics on spaces of constant negative curvature is essentially a number-theoretic problem. *A posteriori*, one can say that zeta functions arise in both classical and quantum mechanics because in both 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 = tr log*, and in this way the spectrum of an evolution operator becames related to its traces, i.e., periodic orbits. A perhaps deeper way of restating this is to observe that the trace formulas perform the same service in all of the above problems; they relate the spectrum of lengths (local dynamics) to the spectrum of eigenvalues (global averages), and for nonlinear geometries they play a role analogous to that the Fourier transform plays for the circle.

[section 17.1]

[exercise 4.1]

In M. Gutzwiller words:

> "The classical periodic orbits are a crucial stepping stone in the understanding of quantum mechanics, in particular when then 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. The mathematical proof, however, is based on the high degree of symmetry of these surfaces which can be compared to the sphere, although the negative curvature allows for many more different shapes."

## A.4 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*

In an afternoon of 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 . . .

34 years old at the time, Dieter was a scruffy kind of guy, always in sandals and holed out jeans, a left winger and a mountain climber, working around the clock with his students Gregor and Klaus to complete the work that Bohr himself would have loved to see 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 classical 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 Kramers to work on the problem, and 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. In 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:

> (...) 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, and 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 classical 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 [5] added the topological indices in 1980, and in 1991 Wintgen and collaborators [8, 9] 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 classical 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!"

### A.4.1 Berry-Keating conjecture

A very appealing proposal in the context of semiclassical quantization is due to M. Berry and J. Keating [21]. 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 [2, 1] described above, but the philosophy is quite different; the claim is that the optimal estimate for low eigenvalues of classically 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 maybe not in its final form yet, it is very intriguing and worth pursuing.

The real life challenge are generic dynamical flows, which fit neither of the above two idealized settings.

## Commentary

**Remark A.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 A.2** 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 apporach the set by some given $\epsilon^n$, 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 (geometry of it matters not at all), take an $\epsilon$ neighborhood around it. Given that the Lebesgue measure is $\epsilon^\alpha$ 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 [32]. A more readable proof is given in M. Benedicks and L.-S. Young [13].

(based on Ya.B. Pesin's comments)

**Remark A.3** Einstein did it? The first hint that chaos is afoot in quantum mechanics was given in a note by A. Einstein [16]. 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 [17, 18]. We asked about the paper two people from that era, Sir Rudolf Peierls and Abraham Pais, and both knew nothing about 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 Méchanique Céleste [19]. In 1954 Fritz Reiche, who had previously followed Einstein as professor of physics in Wroclaw (??), pointed out to J.B. Keller that Keller's geometrical semiclassical quantization was anticipated by the long forgotten paper by A. Einstein [16]. 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 [19], 41 years later. But before Ian Percival included the topological phase, and Wintgen and students recycled the Helium atom, knowing Méchanique Céleste was not enough to complete Bohr's original program.

**Remark A.4** Sources. The tale of appendix A.4, aside from a few personal recollections, is in large part lifted from Abraham Pais' accounts of the demise of the old quantum theory [6, 7], as well as Jammer's account [2]. In August 1994 Dieter Wintgen died in a climbing accident in the Swiss Alps.

## References

[A.1] F. Diacu and P. Holmes, *Celestial Encounters, The Origins of Chaos and Stability* (Princeton Univ. Press, Princeton NJ 1996).

[A.2] M. Jammer, *The Conceptual Development of Quantum mechanics* (McGraw-Hill, New York 1966).

[A.3] J. Mehra and H. Rechtenberg, *The Historical Development of the Quantum Theory* (Springer, New York 1982).

[A.4] M. Born, *Vorlesungen über Atommechanik* (Springer, Berlin 1925). English translation: *The Mechanics of the Atom*, (F. Ungar Publishing Co., New York 1927).

[A.5] J. G. Leopold and I. Percival, *J. Phys. B*, **13**, 1037 (1980).

[A.6] A. Pais, *Inward Bound: of Matter and Forces in the Physical World* (Oxford Univ. Press, Oxford 1986).

[A.7] A. Pais, *Niels Bohr's Times, in Physics, Philosophy and Polity* (Oxford Univ. Press, Oxford 1991).

[A.8] G.S. Ezra, K. Richter, G. Tanner and D. Wintgen, " Semiclassical cycle expansion for the helium atom," *J. Phys. B* **24**, L413 (1991).

[A.9] D. Wintgen, K. Richter and G. Tanner, " The semiclassical helium atom," *CHAOS* **2**, 19 (1992).

[A.10] E. Hopf, "Abzweigung einer periodischen Lösung," *Bereich. Sächs. Acad. Wiss. Leipzig, Math. Phys. Kl.* **94**, 15 (1942); "Bifurcation of a periodic solution from a stationary solution of a system of differential equations," transl. by L. N. Howard and N. Kopell, in *The Hopf bifurcation and its applications*, J. E. Marsden and M. McCracken, eds., pp. 163-193, (Springer-Verlag, New York 1976).

[A.11] E. Hopf, "A mathematical example displaying features of turbulence," *Commun. Appl. Math.* **1**, 303 (1948).

[A.12] D.W. Moore and E.A. Spiegel, "A thermally excited nonlinear oscillator," *Astrophys. J.*, **143**, 871 (1966).

[A.13] N.H. Baker, D.W. Moore and E.A. Spiegel, *Quar. J. Mech. and Appl. Math.* **24**, 391 (1971).

[A.14] E.A. Spiegel, *Chaos: a mixed metaphor for turbulence*, *Proc. Roy. Soc.* **A413**, 87 (1987).

[A.15] M.L. Cartwright and J.E. Littlewood, "On nonlinear differential equations of the second order,"

[A.16] A. Einstein, "On the Quantum Theorem of Sommerfeld and Epstein," p. 443, English translation of "Zum Quantensatz von Sommerfeld und Epstein," *Verh. Deutsch. Phys. Ges.* **19**, 82 (1917), in *The Collected Papers of Albert Einstein*, Volume **6**: *The Berlin Years: Writings, 1914-1917*, A. Engel, transl. and E. Schucking, (Princeton University Press, Princeton, New Jersey 1997).

[A.17] M.C. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer, New York 1990).

[A.18] D. Stone, "1917 Einstein paper," *Physics Today* (15 Sep 2005)

[A.19] J.B. Keller, "Corrected Bohr-Sommerfeld quantum conditions for nonseparable systems," *Ann. Phys. (N.Y.)* **4**, 180 (1958).

[A.20] A. Selberg, *J. Ind. Math. Soc.* **20**, 47 (1956).

[A.21] M.V. Berry and J.P. Keating, *J. Phys.* **A 23**, 4839 (1990).

# Appendix B

# 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 punch line (B.22):

Hamilton-Cayley equation $\prod(\mathbf{M} - \lambda_i \mathbf{1}) = 0$ associates with each distinct root $\lambda_i$ of a matrix $\mathbf{M}$ a projection onto $i$th vector subspace

$$\mathbf{P}_i = \prod_{j \neq i} \frac{\mathbf{M} - \lambda_j \mathbf{1}}{\lambda_i - \lambda_j} \,.$$

## B.1 Linear algebra

The reader might prefer going straight to sect. B.2.

**Vector space.** A set $V$ of elements $\mathbf{x}, \mathbf{y}, \mathbf{z}, \ldots$ 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

$$\begin{aligned} a(\mathbf{x}+\mathbf{y}) &= a\mathbf{x} + a\mathbf{y}\,, \qquad a, b \in \mathbb{F}\,, \quad \mathbf{x}, \mathbf{y} \in V \\ (a+b)\mathbf{x} &= a\mathbf{x} + b\mathbf{x} \\ a(b\mathbf{x}) &= (ab)\mathbf{x} \\ 1\,\mathbf{x} &= \mathbf{x}\,, \qquad 0\,\mathbf{x} = \mathbf{0}\,. \end{aligned} \tag{B.1}$$

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.** $\{\mathbf{e}^{(1)}, \cdots, \mathbf{e}^{(d)}\}$ is any linearly independent subset of $V$ whose span is $V$. The number of basis elements $d$ is the *dimension* of the vector space $V$.

**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* $\mathbf{g}^\dagger = (\mathbf{g}^{-1})^T$ is the transpose of the inverse of $\mathbf{g}$. This observation motivates introduction of a *dual* representation space $\overline{V}$, 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* $\overline{V}$ is the set of all linear forms on $V$ over the field $\mathbb{F}$.

If $\{\mathbf{e}^{(1)}, \cdots, \mathbf{e}^{(d)}\}$ is a basis of $V$, then $\overline{V}$ is spanned by the *dual basis* $\{\mathbf{e}_{(1)}, \cdots, \mathbf{e}_{(d)}\}$, the set of $d$ linear forms $\mathbf{e}_{(k)}$ such that

$$\mathbf{e}_{(j)} \cdot \mathbf{e}^{(k)} = \delta^k_j\,,$$

where $\delta^k_j$ is the Kronecker symbol, $\delta^k_j = 1$ if $j = k$, and zero otherwise. The components of dual representation space vectors $\overline{y} \in \overline{V}$ will here be distinguished by upper indices

$$(y^1, y^2, \ldots, y^n)\,. \tag{B.2}$$

They transform under $GL(n, \mathbb{F})$ as

$$y'^a = (\mathbf{g}^\dagger)^a{}_b y^b\,. \tag{B.3}$$

For $GL(n, \mathbb{F})$ no complex conjugation is implied by the $^\dagger$ notation; that interpretation applies only to unitary subgroups $U(n) \subset GL(n, \mathbb{C})$. $\mathbf{g}$ can be distinguished from $\mathbf{g}^\dagger$ by meticulously keeping track of the relative ordering of the indices,

$$(\mathbf{g})^b_a \to g_a{}^b\,, \qquad (\mathbf{g}^\dagger)^b_a \to g^b{}_a\,. \tag{B.4}$$

**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 $\mathcal{T} \cdot \mathcal{T} \to \mathcal{T}$, 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 $\mathcal{T}$:

$$\mathbf{t}_\alpha \cdot \mathbf{t}_\beta = \sum_{\gamma=0}^{r-1} \tau_{\alpha\beta}{}^\gamma \mathbf{t}_\gamma \,, \qquad \tau_{\alpha\beta}{}^\gamma \in \mathbb{C}\,; \tag{B.5}$$

(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,

$$(\mathbf{t}_\alpha)_\beta{}^\gamma \equiv \tau_{\alpha\beta}{}^\gamma \,, \tag{B.6}$$

whose dimension is the dimension 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*

$$(\mathbf{t}_\alpha \cdot \mathbf{t}_\beta)^c_a = (t_\alpha)^b_a (t_\beta)^c_b \,, \qquad \mathbf{t}_\alpha \in V \otimes \bar{V} \,, \tag{B.7}$$

and the *Lie product*

$$(\mathbf{t}_\alpha \cdot \mathbf{t}_\beta)^c_a = (t_\alpha)^b_a (t_\beta)^c_b - (t_\alpha)^b_c (t_\beta)^a_b \,, \qquad \mathbf{t}_\alpha \in V \otimes \bar{V} \tag{B.8}$$

which defines a *Lie algebra*.

## B.2 Eigenvalues and eigenvectors

Eigenvalues of a $[d \times d]$ matrix $\mathbf{M}$ are the roots of its characteristic polynomial

$$\det(\mathbf{M} - \lambda \mathbf{1}) = \prod (\lambda_i - \lambda) = 0 \,. \tag{B.9}$$

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 $\lambda_i$

$$\mathbf{M}\mathbf{e}^{(i)} = \lambda_i \mathbf{e}^{(i)} \,. \tag{B.10}$$

If $\lambda_i \neq \lambda_j$, $\mathbf{e}^{(i)}$ and $\mathbf{e}^{(j)}$ are linearly independent, so there are at most $d$ distinct eigenvalues, which we assume have been computed by some method, and ordered by their real parts, $\mathrm{Re}\,\lambda_\mathrm{i} \geq \mathrm{Re}\,\lambda_{\mathrm{i}+1}$.

**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 $\mathbf{x} \in \mathbb{R}^d$

$$\mathbf{x} = x_1\,\mathbf{e}^{(1)} + x_2\,\mathbf{e}^{(2)} + \cdots + x_d\,\mathbf{e}^{(d)}\,. \tag{B.11}$$

From (B.10) it follows that matrix $(\mathbf{M} - \lambda_i\mathbf{1})$ annihilates $\mathbf{e}^{(i)}$,

$$(\mathbf{M} - \lambda_i\mathbf{1})\mathbf{e}^{(j)} = (\lambda_j - \lambda_i)\mathbf{e}^{(j)}\,,$$

and the product of all such factors annihilates any vector, so the matrix $\mathbf{M}$ satisfies its characteristic equation (B.9),

$$\prod_{i=1}^{d}(\mathbf{M} - \lambda_i\mathbf{1}) = 0\,. \tag{B.12}$$

This humble fact has a name: the Hamilton-Cayley theorem. If we delete one term from this product, we find that the remainder projects $\mathbf{x}$ onto the corresponding eigenvector:

$$\prod_{j\neq i}(\mathbf{M} - \lambda_j\mathbf{1})\mathbf{x} = \prod_{j\neq i}(\lambda_i - \lambda_j)x_i\mathbf{e}^{(i)}\,.$$

Dividing through by the $(\lambda_i - \lambda_j)$ factors yields the *projection operators*

$$\mathbf{P}_i = \prod_{j\neq i}\frac{\mathbf{M} - \lambda_j\mathbf{1}}{\lambda_i - \lambda_j}\,, \tag{B.13}$$

which are *orthogonal* and *complete*:

$$\mathbf{P}_i\mathbf{P}_j = \delta_{ij}\mathbf{P}_j\,, \quad \text{(no sum on } j)\,, \qquad \sum_{i=1}^{r}\mathbf{P}_i = \mathbf{1}\,. \tag{B.14}$$

By (B.10) every column of $\mathbf{P}_i$ 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 (B.14), they are mutually orthogonal,

$$\mathbf{e}_{(i)}\cdot\mathbf{e}^{(j)} = c\,\delta_i^j\,. \tag{B.15}$$

The non-zero constant $c$ is convention dependent and not worth fixing, unless you feel nostalgic about Clebsch-Gordan coefficients. It follows from the characteristic equation (B.12) that $\lambda_i$ is the eigenvalue of $\mathbf{M}$ on $\mathbf{P}_i$ subspace:

$$\mathbf{M}\mathbf{P}_i = \lambda_i\mathbf{P}_i \qquad \text{(no sum on } i)\,. \tag{B.16}$$

Using $\mathbf{M} = \mathbf{M}\,\mathbf{1}$ and completeness relation (B.14) we can rewrite $\mathbf{M}$ as

$$\mathbf{M} = \lambda_1 \mathbf{P}_1 + \lambda_2 \mathbf{P}_2 + \cdots + \lambda_d \mathbf{P}_d \,. \tag{B.17}$$

Any matrix function $f(\mathbf{M})$ takes the scalar value $f(\lambda_i)$ on the $\mathbf{P}_i$ subspace, $f(\mathbf{M})\mathbf{P}_i = f(\lambda_i)\mathbf{P}_i$, and is easily evaluated through its *spectral decomposition*

$$f(\mathbf{M}) = \sum_i f(\lambda_i)\mathbf{P}_i \,. \tag{B.18}$$

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.

**Example B.1** ***Complex eigenvalues.*** *As* $\mathbf{M}$ *has only real entries, it will in general have either real eigenvalues, or complex conjugate pairs of eigenvalues. That is not surprising, but also 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* $\lambda_k, \lambda_{k+1}$ *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,* $\{\mathbf{e}^{(k)}, \mathbf{e}^{(k+1)}\} \to \{\mathrm{Re}\,\mathbf{e}^{(k)}, \mathrm{Im}\,\mathbf{e}^{(k)}\}$*. In this 2-d real representation the block* $\mathbf{M}' \to \mathbf{N}$ *consists of the identity and the generator of* $SO(2)$ *rotations*

$$\mathbf{N} = \begin{pmatrix} \mu & -\omega \\ \omega & \mu \end{pmatrix} = \mu \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \omega \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} .$$

*Trajectories of* $\dot{\mathbf{x}} = \mathbf{N}\,\mathbf{x}$, $\mathbf{x}(t) = J^t\,\mathbf{x}(0)$*, where*

$$J^t = e^{t\mathbf{N}} = e^{t\mu} \begin{pmatrix} \cos\omega t & -\sin\omega t \\ \sin\omega t & \cos\omega t \end{pmatrix} , \tag{B.19}$$

*spiral in/out around* $(x, y) = (0, 0)$*, see figure 4.4, with the rotation period* $T$ *and the expansion/contraction multiplier along the* $\mathbf{e}^{(j)}$ *eigendirection per a turn of the spiral:* [exercise B.1]

$$T = 2\pi/\omega \,, \qquad \Lambda_{radial} = e^{T\mu} \,, \qquad \Lambda_j = e^{T\mu^{(j)}} \,. \tag{B.20}$$

*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* $10^{-2}T$*).* $\Lambda_j$ *multipliers give us estimates of strange-set thickness.*

While for a randomly constructed matrix all eigenvalues are distinct with probability 1, that is not true in presence of symmetries. What can one say about situation where $d_\alpha$ eigenvalues are degenerate, $\lambda_\alpha = \lambda_i = \lambda_{i+1} = \cdots = \lambda_{i+d_\alpha-1}$? Hamilton-Cayley (B.12) now takes form

$$\prod_{\alpha=1}^{r} (\mathbf{M} - \lambda_\alpha \mathbf{1})^{d_\alpha} = 0 \,, \qquad \sum_\alpha d_\alpha = d \,. \tag{B.21}$$

We distinguish two cases:

**M can be brought to diagonal form.** The characteristic equation (B.21) can be replaced by the minimal polynomial,

$$\prod_{\alpha=1}^{r}(\mathbf{M}-\lambda_\alpha\mathbf{1}) = 0\,, \tag{B.22}$$

where the product includes each distinct eigenvalue only once. Matrix $\mathbf{M}$ satisfies

$$\mathbf{M}\mathbf{e}^{(\alpha,k)} = \lambda_i\mathbf{e}^{(\alpha,k)}\,, \tag{B.23}$$

on a $d_\alpha$-dimensional subspace spanned by a linearly independent set of basis eigenvectors $\{\mathbf{e}^{(\alpha,1)},\mathbf{e}^{(\alpha,2)},\cdots,\mathbf{e}^{(\alpha,d_\alpha)}\}$. This is the easy case whose discussion we continue in appendix H.2.1. Luckily, if the degeneracy is due to a finite or compact symmetry group, relevant $\mathbf{M}$ matrices can always be brought to such Hermitian, 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 B.2.

**Example B.2** ***Decomposition of 2-d 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* $[2\times2]$ *non-singular matrix*

$$\mathbf{M} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}.$$

*takes us a long way toward developing intuition about arbitrary finite-dimensional matrices. The eigenvalues*

$$\lambda_{1,2} = \frac{1}{2}\mathrm{tr}\,\mathbf{M} \pm \frac{1}{2}\sqrt{(\mathrm{tr}\,\mathbf{M})^2 - 4\det\mathbf{M}} \tag{B.24}$$

*are the roots of the characteristic (secular) equation*

$$\begin{aligned}\det(\mathbf{M}-\lambda\,\mathbf{1}) &= (\lambda_1-\lambda)(\lambda_2-\lambda)\\ &= \lambda^2 - \mathrm{tr}\,\mathbf{M}\,\lambda + \det\mathbf{M} \;=\; 0\,.\end{aligned}$$

**Distinct eigenvalues case** *has already been described in full generality. The left/right eigenvectors are the rows/columns of projection operators*

$$P_1 = \frac{\mathbf{M}-\lambda_2\mathbf{1}}{\lambda_1-\lambda_2}\,,\qquad P_2 = \frac{\mathbf{M}-\lambda_1\mathbf{1}}{\lambda_2-\lambda_1}\,,\qquad \lambda_1\neq\lambda_2\,. \tag{B.25}$$

**Degenerate eigenvalues.** *If $\lambda_1 = \lambda_2 = \lambda$, we distinguish two cases: (a) $\mathbf{M}$ can be brought to diagonal form. This is the easy case whose discussion in any dimension we continue in appendix H.2.1. (b) $\mathbf{M}$ 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*

$$\mathbf{M} = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}, \qquad \mathbf{e}^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}^{(2)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

*$\mathbf{v}^{(2)}$ helps span the 2-d space, $(\mathbf{M} - \lambda)^2\mathbf{v}^{(2)} = 0$, but is not an eigenvector, as $\mathbf{M}\mathbf{v}^{(2)} = \lambda\mathbf{v}^{(2)} + \mathbf{e}^{(1)}$. For every such Jordan $[d_\alpha \times d_\alpha]$ block there is only one eigenvector per block. Noting that*

$$\mathbf{M}^m = \begin{pmatrix} \lambda^m & m\lambda^{m-1} \\ 0 & \lambda^m \end{pmatrix},$$

*we see that instead of acting multiplicatively on $\mathbb{R}^2$, fundamental matrix $J^t = \exp(t\mathbf{M})$*

$$e^{t\mathbf{M}}\begin{pmatrix} u \\ v \end{pmatrix} = e^{t\lambda}\begin{pmatrix} u + tv \\ v \end{pmatrix} \tag{B.26}$$

*picks up a power-low correction. That spells trouble (logarithmic term $\ln t$ if we bring the extra term into the exponent).*

**Example B.3** ***Projection operator decomposition in 2-$d$:*** *Let's illustrate how the distinct eigenvalues case works with the $[2\times 2]$ matrix*

$$\mathbf{M} = \begin{pmatrix} 4 & 1 \\ 3 & 2 \end{pmatrix}.$$

*Its eigenvalues $\{\lambda_1, \lambda_2\} = \{5, 1\}$ are the roots of (B.24):*

$$\det(\mathbf{M} - \lambda\mathbf{1}) = \lambda^2 - 6\lambda + 5 = (5 - \lambda)(1 - \lambda) = 0\,.$$

*That $\mathbf{M}$ satisfies its secular equation (Hamilton-Cayley theorem) can be verified by explicit calculation:*

$$\begin{pmatrix} 4 & 1 \\ 3 & 2 \end{pmatrix}^2 - 6\begin{pmatrix} 4 & 1 \\ 3 & 2 \end{pmatrix} + 5\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

*Associated with each root $\lambda_i$ is the* projection operator *(B.25)*

$$P_1 = \frac{1}{4}(\mathbf{M} - \mathbf{1}) = \frac{1}{4}\begin{pmatrix} 3 & 1 \\ 3 & 1 \end{pmatrix} \tag{B.27}$$

$$P_2 = \frac{1}{4}(\mathbf{M} - 5\cdot\mathbf{1}) = \frac{1}{4}\begin{pmatrix} 1 & -1 \\ -3 & 3 \end{pmatrix}. \tag{B.28}$$

*Matrices $\mathbf{P}_i$ are orthonormal and complete, The dimension of the $i$th subspace is given by $d_i = \operatorname{tr}\mathbf{P}_i$; in case at hand both subspaces are 1-dimensional. From the characteristic equation it follows that $\mathbf{P}_i$ satisfies the eigenvalue equation $\mathbf{M}\,\mathbf{P}_i = \lambda_i\mathbf{P}_i$. Two consequences are immediate. First, we can easily evaluate any function of $\mathbf{M}$ by spectral decomposition*

$$\mathbf{M}^7 - 3\cdot\mathbf{1} = (5^7 - 3)\mathbf{P}_1 + (1 - 3)\mathbf{P}_2 = \begin{pmatrix} 58591 & 19531 \\ 58593 & 19529 \end{pmatrix}.$$

*Second, as $\mathbf{P}_i$ 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:*

$$\begin{aligned}\{\mathbf{e}^{(1)}, \mathbf{e}^{(2)}\} &= \{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\-3\end{pmatrix}\}\\ \{\mathbf{e}_{(1)}, \mathbf{e}_{(2)}\} &= \{(3 \quad 1), (1 \quad -1)\},\end{aligned}$$

*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 orthonormal (B.15) by inspection.*

## B.3 Stability of Hamiltonian flows

(M.J. Feigenbaum and P. Cvitanović)

The symplectic structure of Hamilton's equations buys us much more than the incompressibility, or the phase space volume conservation alluded to in sect.7.1. The evolution equations for any $p, q$ dependent quantity $Q = Q(q, p)$ are given by (14.32).

In terms of the Poisson brackets, the time evolution equation for $Q = Q(q, p)$ is given by (14.34). We now recast the symplectic condition (7.11) in a form convenient for using the symplectic constraints on $M$. Writing $x(t) = x' = [p', q']$ and the fundamental matrix and its inverse

$$M = \begin{pmatrix} \frac{\partial q'}{\partial q} & \frac{\partial q'}{\partial p} \\ \frac{\partial p'}{\partial q} & \frac{\partial p'}{\partial p} \end{pmatrix}, \qquad M^{-1} = \begin{pmatrix} \frac{\partial q}{\partial q'} & \frac{\partial q}{\partial p'} \\ \frac{\partial p}{\partial q'} & \frac{\partial p}{\partial p'} \end{pmatrix}, \tag{B.29}$$

we can spell out the symplectic invariance condition (7.11):

$$\begin{aligned}\frac{\partial q'_k}{\partial q_i}\frac{\partial p'_k}{\partial q_j} - \frac{\partial p'_k}{\partial q_i}\frac{\partial q'_k}{\partial q_j} &= 0\\ \frac{\partial q'_k}{\partial p_i}\frac{\partial p'_k}{\partial p_j} - \frac{\partial p'_k}{\partial p_i}\frac{\partial q'_k}{\partial p_j} &= 0\\ \frac{\partial q'_k}{\partial q_i}\frac{\partial p'_k}{\partial p_j} - \frac{\partial p'_k}{\partial q_i}\frac{\partial q'_k}{\partial p_j} &= \delta_{ij}\,.\end{aligned} \tag{B.30}$$

From (7.18) we obtain

$$\frac{\partial q_i}{\partial q'_j} = \frac{\partial p'_j}{\partial p_i}, \quad \frac{\partial p_i}{\partial p'_j} = \frac{\partial q'_j}{\partial q_i}, \quad \frac{\partial q_i}{\partial p'_j} = -\frac{\partial q'_j}{\partial p_i}, \quad \frac{\partial p_i}{\partial q'_j} = -\frac{\partial p'_j}{\partial q_i}\,. \tag{B.31}$$

Taken together, (B.31) and (B.30) imply that the flow conserves the $\{p, q\}$ Poisson brackets

$$\begin{aligned}\{q_i, q_j\} &= \frac{\partial q_i}{\partial p'_k}\frac{\partial q_j}{\partial q'_k} - \frac{\partial q_j}{\partial p'_k}\frac{\partial q_i}{\partial q'_k} = 0\\ \{p_i, p_j\} &= 0\,, \qquad \{p_i, q_j\} = \delta_{ij}\,,\end{aligned} \tag{B.32}$$

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 $(2D)^2 - 2D(D-1)/2 - D^2 = 2D^2 + D$ elements of $M$ are linearly independent, as it behooves group elements of the symplectic group $Sp(2D)$.

## B.4 Monodromy matrix for Hamiltonian flows

(G. Tanner)

It is not the fundamental matrix of the flow, but the *monodromy* matrix, 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 fundamental matrix $M$. 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 $\delta t$ time independent. The projection of any displacement on $\delta x$ 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 $\mathbf{U}(x(t))$:

$$\tilde{M}(x(t)) = \mathbf{U}^{-1}(x(t))\, M(x(t))\, \mathbf{U}(x(0)) \tag{B.33}$$

These lead to

$$\begin{aligned} \dot{\tilde{M}} &= \tilde{\mathbf{L}}\,\tilde{M} \\ \text{with} \qquad \tilde{\mathbf{L}} &= \mathbf{U}^{-1}(\mathbf{L}\mathbf{U} - \dot{\mathbf{U}}) \end{aligned} \tag{B.34}$$

Note that the properties a) – c) are only fulfilled for $\tilde{M}$ 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 (B.33) at any time $t$. Setting $\mathbf{U} = (x_t^T, x_E^T, x_1^T, \ldots, x_{2d-2}^T)$ gives

$$\tilde{M} = \begin{pmatrix} 1 & * & * & \ldots & * \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & * & & & \\ \vdots & \vdots & & \mathbf{m} & \\ 0 & * & & & \end{pmatrix}; \qquad \tilde{\mathbf{L}} = \begin{pmatrix} 0 & * & * & \ldots & * \\ 0 & 0 & 0 & \ldots & 0 \\ 0 & * & & & \\ \vdots & \vdots & & \mathbf{l} & \\ 0 & * & & & \end{pmatrix}, \tag{B.35}$$

The matrix $\mathbf{m}$ is now the monodromy matrix and the equation of motion are given by

$$\dot{\mathbf{m}} = \mathbf{l}\,\mathbf{m}. \tag{B.36}$$

The vectors $x_1, \ldots, x_{2d-2}$ 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.,

$$\mathbf{U}(t) = (x_t, x_1, x_E, x_2) = \begin{pmatrix} \dot{x} & -\dot{y} & -\dot{u}/q^2 & -\dot{v}/q^2 \\ \dot{y} & \dot{x} & -\dot{v}/q^2 & \dot{u}/q^2 \\ \dot{u} & \dot{v} & \dot{x}/q^2 & -\dot{y}/q^2 \\ \dot{v} & -\dot{u} & \dot{y}/q^2 & \dot{x}/q^2 \end{pmatrix} \tag{B.37}$$

with $x^T = (x, y; u, v)$ 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 $\mathbf{l}$ are given (B.39). 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 $\Lambda = e^{\pm\lambda}$, $\Lambda = -e^{\pm\lambda}$; $\lambda > 0$ is called the Lyapunov exponent of the periodic orbit.
- *loxodromic*, if $\Lambda = e^{\pm u \pm i\Psi}$ with $u$ and $\Psi$ real. This is the most general case possible only in systems with 3 or more degree of freedoms.

For 2 degrees of freedom, i.e., $\mathbf{m}$ is a [2×2] matrix, the eigenvalues are determined by

$$\lambda = \frac{\mathrm{Tr}(\mathbf{m}) \pm \sqrt{\mathrm{Tr}(\mathbf{m})^2 - 4}}{2}, \tag{B.38}$$

i.e., $\mathrm{Tr}(\mathbf{m}) = 2$ separates stable and unstable behavior.

The $\mathbf{l}$ matrix elements for the local transformation (B.37) are

$$\begin{aligned}
\tilde{\mathbf{l}}_{11} &= \frac{1}{q}[(h_x^2 - h_y^2 - h_u^2 + h_v^2)(h_{xu} - h_{yv}) + 2(h_x h_y - h_u h_v)(h_{xv} + h_{yu}) \\
&\quad -(h_x h_u + h_y h_v)(h_{xx} + h_{yy} - h_{uu} - h_{vv})] \\
\tilde{\mathbf{l}}_{12} &= \frac{1}{q^2}[(h_x^2 + h_v^2)(h_{yy} + h_{uu}) + (h_y^2 + h_u^2)(h_{xx} + h_{vv}) \\
&\quad -2(h_x h_u + h_y h_v)(h_{xu} + h_{yv}) - 2(h_x h_y - h_u h_v)(h_{xy} - h_{uv})] \\
\tilde{\mathbf{l}}_{21} &= -(h_x^2 + h_y^2)(h_{uu} + h_{vv}) - (h_u^2 + h_v^2)(h_{xx} + h_{yy}) \\
&\quad +2(h_x h_u - h_y h_v)(h_{xu} - h_{yv}) + 2(h_x h_v + h_y h_u)(h_{xv} + h_{yu}) \\
\tilde{\mathbf{l}}_{22} &= -\tilde{\mathbf{l}}_{11},
\end{aligned} \tag{B.39}$$

with $h_i, h_{ij}$ is the derivative of the Hamiltonian $H$ with respect to the phase space coordinates and $q = |\nabla H|^2$.

## Exercises

B.1. **Real representation of complex eigenvalues.** (Verification of example B.1.) $\lambda_k, \lambda_{k+1}$ 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,

$$\mathbf{P} = \mathbf{P}_k\,, \qquad \mathbf{P}^* = \mathbf{P}_{k+1}\,,$$

where we denote $\mathbf{P}_k$ by $\mathbf{P}$ for notational brevity.

(b) $\mathbf{P}$ can be written as

$$\mathbf{P} = \frac{1}{2}(\mathbf{R} + i\mathbf{Q})\,,$$

where $\mathbf{R} = \mathbf{P}_k + \mathbf{P}_{k+1}$ and $\mathbf{Q}$ are matrices with real elements.

(c) $$\begin{pmatrix} \mathbf{P}_k \\ \mathbf{P}_{k+1} \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}\begin{pmatrix} \mathbf{R} \\ \mathbf{Q} \end{pmatrix}.$$

(d) $\cdots + \lambda_k \mathbf{P}_k + \lambda_k^* \mathbf{P}_{k+1} + \cdots$ complex eigenvalue pair in the spectral decomposition (B.17) is now replaced by a real $[2\times 2]$ matrix

$$\cdots + \begin{pmatrix} \mu & -\omega \\ \omega & \mu \end{pmatrix}\begin{pmatrix} \mathbf{R} \\ \mathbf{Q} \end{pmatrix} + \cdots$$

or whatever is the clearest way to write this real representation.

(P. Cvitanović)

# Appendix C

# Implementing evolution

## C.1 Koopmania

THE WAY in which time evolution acts on observables may be rephrased in the language of functional analysis, by introducing the *Koopman operator*, whose action on a state space function $a(x)$ is to replace it by its downstream value time $t$ later, $a(x) \to a(x(t))$ evaluated at the trajectory point $x(t)$:

$$\mathcal{K}^t a(x) \;=\; a(f^t(x))\,. \tag{C.1}$$

Observable $a(x)$ has no explicit time dependence; all the time dependence comes from its evaluation at $x(t)$ rather than at $x = x(0)$.

Suppose we are starting with an initial density of representative points $\rho(x)$: then the average value of $a(x)$ evolves as

$$\langle a\rangle(t) \;=\; \frac{1}{|\rho_{\mathcal{M}}|}\int_{\mathcal{M}} dx\, a(f^t(x))\rho(x) \;=\; \frac{1}{|\rho_{\mathcal{M}}|}\int_{\mathcal{M}} dx\, \left[\mathcal{K}^t a(x)\right]\rho(x)\,.$$

An alternative point of view (analogous to the shift from the Heisenberg to the Schrödinger picture in quantum mechanics) 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 (14.10) depends on the trajectory point time $t$ in the past, so the Perron-Frobenius operator is the adjoint of the Koopman operator

$$\int_{\mathcal{M}} dx\, \left[\mathcal{K}^t a(x)\right]\rho(x) = \int_{\mathcal{M}} dx\, a(x)\left[\mathcal{L}^t\rho(x)\right]\,. \tag{C.2}$$

Checking this is an easy change of variables exercise. For finite dimensional deterministic invertible flows the Koopman operator (C.1) is simply the inverse of

the Perron-Frobenius operator (14.6), so in what follows we shall not distinguish the two. However, for infinite dimensional flows contracting forward in time and for stochastic flows such inverses do not exist, and there you need to be more careful.

The family of Koopman's operators $\{\mathcal{K}^t\}_{t\in\mathbb{R}_+}$ forms a semigroup parameterized by time

(a) $\mathcal{K}^0 = \mathbf{1}$

(b) $\mathcal{K}^t\mathcal{K}^{t'} = \mathcal{K}^{t+t'} \qquad t, t' \geq 0 \qquad$ (semigroup property) ,

with the *generator* of the semigroup, the generator of infinitesimal time translations defined by

$$\mathcal{A} = \lim_{t\to 0^+} \frac{1}{t}\left(\mathcal{K}^t - \mathbf{1}\right) .$$

(If the flow is finite-dimensional and invertible, $\mathcal{A}$ is a generator of a group). The explicit form of $\mathcal{A}$ follows from expanding dynamical evolution up to first order, as in (2.5):

$$\mathcal{A}a(x) = \lim_{t\to 0^+} \frac{1}{t}\left(a(f^t(x)) - a(x)\right) = v_i(x)\partial_i a(x) . \tag{C.3}$$

Of course, that is nothing but the definition of the time derivative, so the equation of motion for $a(x)$ is

$$\left(\frac{d}{dt} - \mathcal{A}\right) a(x) = 0 . \tag{C.4}$$

[appendix C.2]

The finite time Koopman operator (C.1) can be formally expressed by exponentiating the time evolution generator $\mathcal{A}$ as

$$\mathcal{K}^t = e^{t\mathcal{A}} . \tag{C.5}$$

[exercise C.1]

The generator $\mathcal{A}$ looks very much like the generator of translations. Indeed, for a constant velocity field dynamical evolution is nothing but a translation by time $\times$ velocity:

[exercise 14.10]

$$e^{tv\frac{\partial}{\partial x}} a(x) = a(x + tv) . \tag{C.6}$$

As we will not need to implement a computational formula for general $e^{t\mathcal{A}}$ in what follows, we relegate making sense of such operators to appendix C.2. Here we limit ourselves to a brief remark about the notion of "spectrum" of a linear operator.

[appendix C.2]

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 $\|\mathcal{K}^t\| \leq Me^{t\beta}$ 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. J.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 (15.20). 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

$$\int_0^\infty dt\, e^{-st}\mathcal{K}^t = \frac{1}{s-\mathcal{A}}\,, \qquad \mathrm{Re}\, s > \beta\,, \tag{C.7}$$

that the *resolvent* operator $(s-\mathcal{A})^{-1}$ is bounded ("resolvent" = able to cause separation into constituents) [section J.2]

$$\left\|\frac{1}{s-\mathcal{A}}\right\| \leq \int_0^\infty dt\, e^{-st} M e^{t\beta} = \frac{M}{s-\beta}\,.$$

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 (14.29) into the resolvent form (14.31) and enables us to study its spectrum.

## C.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 (14.27) as

$$\mathcal{A} = \frac{d}{dt}\mathcal{K}^t\Big|_{t=0}\,.$$

If we now take the derivative at arbitrary times we get

$$\begin{aligned}\left(\frac{d}{dt}\mathcal{K}^t\psi\right)(x) &= \lim_{\eta\to 0}\frac{\psi(f^{t+\eta}(x)) - \psi(f^t(x))}{\eta} \\ &= v_i(f^t(x))\,\frac{\partial}{\partial\tilde{x}_i}\psi(\tilde{x})\Big|_{\tilde{x}=f^t(x)} \\ &= \left(\mathcal{K}^t\mathcal{A}\psi\right)(x)\end{aligned}$$

which can be formally integrated like an ordinary differential equation yielding [exercise C.1]

$$\mathcal{K}^t = e^{t\mathcal{A}} . \tag{C.8}$$

This guarantees that the Laplace transform manipulations in sect. 14.5 are correct. Though the formal expression of the semigroup (C.8) is quite simple one has to take care in implementing its action. If we express the exponential through the power series

$$\mathcal{K}^t = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathcal{A}^k , \tag{C.9}$$

we encounter the problem that the infinitesimal generator (14.27) 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. 4.2 and 4.3 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

$$\mathcal{K}^t = \prod_{m=1}^{t/\delta\tau} (1 + \delta\tau \mathcal{A}_{(m)}) + O(\delta\tau^2) \tag{C.10}$$

where

$$(\mathcal{A}_{(m)}\psi)(x) = v_i(f^{m\delta\tau}(x)) \left. \frac{\partial \psi}{\partial \tilde{x}_i} \right|_{\tilde{x}=f^{m\delta\tau}(x)}$$

As far as the $x$ dependence is concerned, $e^{\delta\tau\mathcal{A}_i}$ acts as

$$e^{\delta\tau\mathcal{A}_i} \left\{ \begin{array}{c} x_1 \\ \cdot \\ x_i \\ x_d \end{array} \right\} \to \left\{ \begin{array}{c} x_1 \\ \cdot \\ x_i + \delta\tau v_i(x) \\ x_d \end{array} \right\} . \tag{C.11}$$

[exercise 2.6]

We see that the product form (C.10) 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 $\delta\tau\times$velocity at each time step.

### C.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 (C.10) by splitting it into infinitesimal steps and keeping terms up to $\delta\tau^2$ in [exercise C.2]

$$\mathcal{K}^{\delta\tau} = \hat{\mathcal{K}}^{\delta\tau} + \frac{1}{24}(\delta\tau)^3[\mathcal{B} + 2\mathcal{C}, [\mathcal{B}, \mathcal{C}]] + \cdots , \tag{C.12}$$

where

$$\hat{\mathcal{K}}^{\delta\tau} = e^{\frac{1}{2}\delta\tau\mathcal{B}} e^{\delta\tau\mathcal{C}} e^{\frac{1}{2}\delta\tau\mathcal{B}} . \tag{C.13}$$

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 (14.30), 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:

$$\begin{aligned}
e^{\frac{1}{2}\delta\tau\mathcal{B}} \begin{Bmatrix} q \\ p \end{Bmatrix} &\rightarrow \begin{Bmatrix} q' \\ p' \end{Bmatrix} = \begin{Bmatrix} q - \frac{\delta\tau}{2} p \\ p \end{Bmatrix} \\
e^{\delta\tau\mathcal{C}} \begin{Bmatrix} q' \\ p' \end{Bmatrix} &\rightarrow \begin{Bmatrix} q'' \\ p'' \end{Bmatrix} = \begin{Bmatrix} q' \\ p' + \delta\tau\partial V(q') \end{Bmatrix} \\
e^{\frac{1}{2}\delta\tau\mathcal{B}} \begin{Bmatrix} q'' \\ p'' \end{Bmatrix} &\rightarrow \begin{Bmatrix} q''' \\ p''' \end{Bmatrix} = \begin{Bmatrix} q' - \frac{\delta\tau}{2} p'' \\ p'' \end{Bmatrix}
\end{aligned} \tag{C.14}$$

Collecting the terms we obtain an integration rule for this type of symplectic flow which is better than the straight Euler integration (C.11) as it is accurate up to order $\delta\tau^2$:

$$\begin{aligned}
q_{n+1} &= q_n - \delta\tau\, p_n - \frac{(\delta\tau)^2}{2} \partial V\,(q_n - \delta\tau p_n/2) \\
p_{n+1} &= p_n + \delta\tau\partial V\,(q_n - \delta\tau p_n/2)
\end{aligned} \tag{C.15}$$

The fundamental matrix of one integration step is given by

$$M = \begin{pmatrix} 1 & -\delta\tau/2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \delta\tau\partial V(q') & 1 \end{pmatrix} \begin{pmatrix} 1 & -\delta\tau/2 \\ 0 & 1 \end{pmatrix} . \tag{C.16}$$

Note that the billiard flow (8.11) 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 C.1** Koopman operators. The "Heisenberg picture" in dynamical systems theory has been introduced by Koopman and Von Neumann [1, 2], see also ref. [8]. Inspired by the contemporary advances in quantum mechanics, Koopman [1] 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 $\exp\left(i\hat{H}t/\hbar\right)$ – the kernel of $\mathcal{L}^t(y,x)$ introduced in (14.16) (see also sect. 15.2) is the analogue of the Green's function discussed here in chapter 30. The relation between the spectrum of the Koopman operator and classical ergodicity was formalized by von Neumann [2]. 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 [8] and Gaspard [9].

**Remark C.2** Symplectic integration. The reviews [7] and [8] offer a good starting point for exploring the symplectic integrators literature. For a higher order integrators of type (C.13), check ref. [13].

## Exercises

C.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 $a(x)$.

C.2. **Non-commutativity.** Check that the commutators in (C.12) are not vanishing by showing that

$$[\mathcal{B}, C] = -p\left(V''\frac{\partial}{\partial p} - V'\frac{\partial}{\partial q}\right).$$

C.3. **Symplectic leapfrog integrator.** Implement (C.15) for 2-dimensional Hamiltonian flows; compare it with Runge-Kutta integrator by integrating trajectories in some (chaotic) Hamiltonian flow.

## References

[C.1] B.O. Koopman, *Proc. Nat. Acad. Sci. USA* **17**, 315 (1931).

[C.2] J. von Neumann, *Ann. Math.* **33**, 587 (1932).

[C.3] B.A. Shadwick, J.C. Bowman, and P.J. Morrison, *Exactly Conservative Integrators*, `chao-dyn/9507012`, Submitted to *SIAM J. Sci. Comput.*

[C.4] D.J.D. Earn, *Symplectic integration without roundoff error*, `astro-ph/9408024`.

[C.5] P. E. Zadunaiski, "On the estimation of errors propagated in the numerical integration of ordinary differential equations," *Numer. Math.* **27**, 21 (1976).

[C.6] K. Feng, "Difference schemes for Hamiltonian formalism and symplectic geometry," *J. Comput. Math.* **4**, 279 (1986).

[C.7] P.J. Channell and C. Scovel, "Symplectic integration of Hamiltonian systems," *Nonlinearity* **3**, 231 (1990).

[C.8] J.M. Sanz-Serna and M.P. Calvo, *Numerical Hamiltonian problems* (Chapman and Hall, London, 1994).

[C.9] J. M. Sanz-Serna, "Geometric integration," pp. 121-143, in *The State of the Art in Numerical Analysis*, I. S. Duff and G. A. Watson, eds., (Clarendon Press, Oxford, 1997).

[C.10] K. W. Morton, "Book Review: Simulating Hamiltonian Dynamics," *SIAM Review* **48**, 621 (2006).

[C.11] B. Leimkuhler and S. Reich, *Simulating Hamiltonian Dynamics* (Cambridge University Press, Cambridge 2005).

[C.12] E. Hairer, Ch. Lubich, and G. Wanner, *Geometric Numerical Integration* (Springer-Verlag, Berlin, 2002).

[C.13] M. Suzuki, "General theory of fractal path integrals with applications to many-body theories and statistical physics," *J. Math. Phys.* **32**, 400 (1991).

# Appendix D

# Symbolic dynamics techniques

THE KNEADING THEORY for unimodal mappings is developed in sect. D.1. The prime factorization for dynamical itineraries of sect. D.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.

## D.1 Topological zeta functions for infinite subshifts

(P. Dahlqvist)

The Markov graph methods outlined in chapter 10 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 $\ldots x_{i-1}x_i x_{i+1} \ldots$ onto a sequence of symbols $\ldots s_{i-1}s_i s_{i+1} \ldots$ where

$$\begin{aligned} s_i &= 0 \quad x_i < x_c \\ s_i &= C \quad x_i = x_c \\ s_i &= 1 \quad x_i > x_c \end{aligned} \tag{D.1}$$

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 $K_\Lambda$ is the itinerary of the critical point. The crucial observation is that no

| $I(C)$ | $\zeta_{top}^{-1}(z)/(1-z)$ | $I(C)$ | $\zeta_{top}^{-1}(z)/(1-z)$ |
|---|---|---|---|
| $1C$ | | $1001C$ | |
| $101C$ | | $100111C$ | |
| $1011101C$ | | $10011C$ | |
| $H^\infty(1)$ | $\prod_{n=0}^{\infty}(1-z^{2^n})$ | $100110C$ | |
| $10111C$ | | $100C$ | |
| $1011111C$ | | $100010C$ | |
| $101^\infty$ | $(1-2z^2)/(1+z)$ | $10001C$ | |
| $10111111C$ | | $100011C$ | |
| $101111C$ | | $1000C$ | |
| $1011C$ | | $100001C$ | |
| $101101C$ | | $10000C$ | |
| $10C$ | $(1-z-z^2)$ | $100000C$ | |
| $10010C$ | | $10^\infty$ | $(1-2z)/(1-z)$ |
| $100101C$ | | | |

**Table D.1:** All ordered kneading sequences up to length seven, as well as some longer kneading sequences. Harmonic extension $H^\infty(1)$ is defined below.

periodic orbit can have a topological coordinate (see sect. D.1.1) 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 orbitsand the same topological zeta function. The topological coordinate of the kneading sequence increases with increasing $\Lambda$.

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*

$$x \mapsto f(x) = \begin{cases} \Lambda x & x \in [0,1/2] \\ \Lambda(1-x) & x \in (1/2,1] \end{cases} , \tag{D.2}$$

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 is periodic or preperiodic are countable and thus of measure zero and consequently *the kneading sequence is aperiodic for almost all* $\Lambda$. 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 $n$, $\underline{K}_\Lambda = PC = s_1 s_2 \ldots s_{n-1} C$ there is a simple expansion for the topological zeta function. Then the expanded zeta function is a polynomial of degree $n$

$$1/\zeta_{\rm top}(z) = \prod_p (1 - z_p^n) = (1 - z) \sum_{i=0}^{n-1} a_i z^i \,, \qquad a_i = \prod_{j=1}^{i} (-1)^{s_j} \tag{D.3}$$

and $a_0 = 1$.

Aperiodic and preperiodic kneading sequences are accounted for by simply replacing $n$ by $\infty$.

**Example**. Consider as an example the kneading sequence $K_\Lambda = 10C$. From (D.3) we get the topological zeta function $1/\zeta_{top}(z) = (1 - z)(1 - z - z^2)$, see table D.1. This can also be realized by redefining the alphabet. The only forbidden subsequence is 100. All allowed periodic orbits, except $\overline{0}$, can can be built from a alphabet with letters $\underline{10}$ and $\underline{1}$. We write this alphabet as $\{\underline{10}, \underline{1}; \overline{0}\}$, yielding the topological zeta function $1/\zeta_{top}(z) = (1 - z)(1 - z - z^2)$. 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 (D.3) we get the topological zeta function $1/\zeta_{top}(z) = (1 - z)(1 - 2z^2)/(1 + z)$, see table D.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 $\{\underline{01^{2n+1}}; \overline{1}, \overline{0}\}$, yielding the topological zeta function above. The pole in the zeta function $\zeta_{top}^{-1}(z)$ is a consequence of the infinite alphabet.

An important consequence of (D.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 $\Lambda$.

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 $\{q\}$ is periodic or not. If the period of the tail is $N$ we can write

$$1/\zeta_{top}(z) = p(z) + q(z)(1 + z^N + z^{2N} \ldots) = p(z) + \frac{q(z)}{1 - z^N} \,,$$

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 [61], 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 13.4. 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 (D.2), for almost all $\Lambda$.

Let us now focus on the relation between the analytic structure of the topological zeta function and the number of periodic orbits, or rather (13.6), the number $N_n$ of fixed points of $f^n(x)$. The trace formula is (see sect. 13.4)

$$N_n = \mathrm{tr}\, T^n = \frac{1}{2\pi i} \oint_{\gamma_r} dz\, z^{-n} \frac{d}{dz} \log \zeta_{top}^{-1}$$

where $\gamma_r$ 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 $\zeta_{top}^{-1}$

$$N_n = \sum_{z_0 : r<|z_0|<R} z_0^{-n} - \sum_{z_p : r<|z_p|<R} z_0^{-n} + \frac{1}{2\pi i} \oint_{\gamma_R} dz\, z^{-n} \frac{d}{dz} \log \zeta_{top}^{-1}$$

and a contribution from a large circle $\gamma_R$. For meromorphic topological zeta functions one may let $R \to \infty$ with vanishing contribution from $\gamma_R$, and $N_n$ will be a sum of exponentials.

The leading zero is associated with the topological entropy, as discussed in chapter 13.

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

$$\Xi(z) = \prod_{n=0}^{\infty} (1 - z^{2^n}) \ . \tag{D.4}$$

This function will appear again when we derive (D.3).

The expansion of $\Xi(z)$ begins as $\Xi(z) = 1 - z - z^2 + z^3 - z^4 + z^5 \ldots$. 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 $\Xi(z) \to 0$ when $z \to \exp(2\pi m/2^n)$ 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 $|\Xi(z)| \to \infty$ when $z \to \exp(2\pi m/3^n)$ 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 $\gamma_R$ 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 $\Lambda$. 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 $\Lambda$ 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 and we get (infinitely) short glimpses of another world beyond the horizon.

We end this section by deriving eqs (D.5) and (D.6). The impenetrable prose is hopefully explained by the accompanying tables.

We know one thing from chapter 10, 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 $a_i$.

| periodic orbits | finite kneading sequences |
|---|---|
| $\overline{P1} = A^\infty(P)$ | |
| | $PC$ |
| $\overline{P0}$ | |
| | $P0PC$ |
| $\overline{P0P1}$ | |
| | $P0P1P0PC$ |
| $\downarrow$ | $\downarrow$ |
| $H^\infty(P)$ | $H^\infty(P)$ |

**Table D.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.

The ordered list of (finite) kneading sequences table D.1 and the ordered list of periodic orbits (on maximal form) are intimately related. In table D.2 we indicate

how they are nested during a period doubling cascade. Every finite kneading sequence $PC$ is bracketed by two periodic orbits, $\overline{P1}$ and $\overline{P0}$. 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 D.2. The periodic orbit $\overline{P1}$ (and the corresponding infinite 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 $PC$ [14] (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.

| $I(C)$ | | | $\zeta_{top}^{-1}(z)/(1-z)$ |
|---|---|---|---|
| $P_1$ | = | $100C$ | $1-z-z^2-z^3$ |
| $H^\infty(P_1)$ | = | $10001001100\ldots$ | $1-z-z^2-z^3-z^4+z^5+z^6+z^7-z^8\ldots$ |
| $P'$ | = | $10001C$ | $1-z-z^2-z^3-z^4+z^5$ |
| $A^\infty(P_2)$ | = | $1000110001\ldots$ | $1-z-z^2-z^3-z^4+z^5-z^6-z^7-z^8\ldots$ |
| $P_2$ | = | $1000C$ | $1-z-z^2-z^3-z^4$ |

**Table D.3:** Example of a step in the iterative construction of the list of kneading sequences $PC$.

Table D.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 D.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

$$\zeta_{P_1}^{-1}(z)\Xi(z^{n_1}) \ , \tag{D.5}$$

and just before $A^\infty(P_2)$

$$\zeta_{P_2}^{-1}(z)/(1-z^{n_2}) \ . \tag{D.6}$$

A short calculation shows that this is exactly what one would obtain by applying (D.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 length 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 (D.5) and (D.6), see Table D.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.

**Remark D.1** How to prove things. 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 proven by combining some theorems of Milnor and Thurston [13]. That approach is hardly instructive in the present context. Our derivation was inspired by Metropolis, Stein and Stein classical paper [14]. For further detail, consult [60].

### D.1.1 Periodic orbits of unimodal maps

A *periodic point* (or a *cycle* point) $x_i$ belonging to a cycle of period $n$ is a real solution of

$$f^n(x_i) = f(f(\ldots f(x_i)\ldots)) = x_i\,, \quad i = 0, 1, 2, \ldots, n-1 \tag{D.7}$$

The $n$th iterate of a unimodal map crosses the diagonal at most $2^n$ times. 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 length $n$. A cycle of length $n$ corresponds to an infinite repetition of a length $n$ symbol string, customarily indicated by a line over the string:

$$S = (s_1 s_2 s_3 \ldots s_n)^\infty = \overline{s_1 s_2 s_3 \ldots s_n}\,.$$

If $\overline{s_1 s_2 \ldots s_n}$ is the symbol string associated with $x_0$, its cyclic permutation $\overline{s_k s_{k+1} \ldots s_n s_1 \ldots s_{k-1}}$ 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 $S'$.

Each cycle yields $n$ rational values of $\gamma$. The repeating string $s_1, s_2, \ldots s_n$ contains an odd number "1"s, the string of well ordered symbols $w_1 w_2 \ldots w_n$ has to be of the double length before it repeats itself. The value $\gamma$ is a geometrical sum which we can write as the finite sum

$$\gamma(\overline{s_1 s_2 \ldots s_n}) = \frac{2^{2n}}{2^{2n}-1} \sum_{t=1}^{2n} w_t/2^t$$

Using this we can calculate the $\hat{\gamma}(S)$ for all short cycles.

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

$$\{a_1, a_2, a_3, a_4, \cdots; \overline{0}\} = \{1, 10, 100, 1000, \ldots; \overline{0}\}\,. \tag{D.8}$$

In this notation the itinerary $S = a_i a_j a_k a_l \cdots$ and the corresponding topological coordinate (**??**) are related by $\gamma(S) = .1^i 0^j 1^k 0^l \cdots$. For example:

$$\begin{array}{lclcl} S & = & 111011101001000\ldots & = & a_1 a_1 a_2 a_1 a_1 a_2 a_3 a_4 \ldots \\ \gamma(S) & = & .101101001110000\ldots & = & .1^1 0^1 1^2 0^1 1^1 0^2 1^3 0^4 \ldots \end{array}$$

Cycle points whose itineraries start with $w_1 = w_2 = \ldots = w_i = 0,\ w_{i+1} = 1$ remain on the left branch of the tent map for $i$ iterations, and satisfy $\gamma(0\ldots 0S) = \gamma(S)/2^i$.

A *periodic point* (or a *cycle* point) $x_i$ belonging to a cycle of period $n$ is a real solution of

$$f^n(x_i) = f(f(\ldots f(x_i)\ldots)) = x_i\,, \quad i = 0, 1, 2, \ldots, n-1\,. \tag{D.9}$$

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 length $n$. A periodic orbit of length $n$ corresponds to an infinite repetition of a length $n$ symbol string, customarily indicated by a line over the string:

$$S = (s_1 s_2 s_3 \ldots s_n)^\infty = \overline{s_1 s_2 s_3 \ldots s_n}\,.$$

As all itineraries are infinite, we shall adopt convention that a finite string itinerary $S = s_1 s_2 s_3 \ldots s_n$ stands for infinite repetition of a finite block, and routinely omit the overline. If $\overline{s_1 s_2 \ldots s_n}$ is the symbol string associated with $x_0$, its cyclic permutation $\overline{s_k s_{k+1} \ldots s_n s_1 \ldots s_{k-1}}$ corresponds to the point $x_{k-1}$ in the same cycle. A periodic orbit $p$ is called *prime* if its itinerary $S$ cannot be written as a repetition of a shorter block $S'$.

Periodic points correspond to rational values of $\gamma$, 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

$$\gamma(a_i a_j \cdots a_k a_\ell) = \begin{cases} \frac{2^n}{2^n-1}.1^i 0^j \cdots 1^k & \text{even} \\ \frac{1}{2^n+1}\left(1 + 2^n \times .1^i 0^j \cdots 1^\ell\right) & \text{odd} \end{cases}\,, \tag{D.10}$$

where $n = i + j + \cdots + k + \ell$ is the cycle period. The maximal value cycle 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:

$$\begin{array}{lclclclcl} \hat{\gamma}(1) & = & \gamma(a_1) & = & .10101\ldots & = & .\overline{10} & = & 2/3 \\ \hat{\gamma}(10) & = & \gamma(a_2) & = & .1^2 0^2 \ldots & = & .\overline{1100} & = & 4/5 \\ \hat{\gamma}(100) & = & \gamma(a_3) & = & .1^3 0^3 \ldots & = & .\overline{111000} & = & 8/9 \\ \hat{\gamma}(101) & = & \gamma(a_2 a_1) & = & .1^2 0^1 \ldots & = & .\overline{110} & = & 6/7 \end{array}$$

An example of a cycle where only the third symbol determines the maximal value cycle point is

$$\hat{\gamma}(1101110) = \gamma(a_2a_1a_2a_1a_1) = .\overline{11011010010010} = 100/129\,.$$

Maximal values of all cycles up to length 5 are given in table!?

## D.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 $\mathbf{P} = \{p_1, p_2, p_3, \cdots\}$ 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 (18.2) 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 .

$$\prod_p \frac{1}{1 - t_p} = \sum_n t_n \tag{D.11}$$

A striking aspect of this formula is its resemblance to the factorization of natural numbers into primes: the relation of the cycle expansion (D.11) to the product over prime cycles is analogous to the Riemann zeta (exercise 17.10) 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: $00 = p_1^2$, $01 = p_1p_2$, $11 = p_2^2$; by ordered concatenations we mean that $p_1p_2$ is legal, but $p_2p_1$ 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 $p_1^{k_1} p_2^{k_2} \cdots p_j^{k_j}$, with primes lexically ordered, we generate the standard list of primes, in agreement with table **??**: 0, 1, 10, 101, 100, 1000, 1001, 1011, 10000, 10001, 10010, 10011, 10110, 10111, 100000, 100001, 100010, 100011, 100110, 100111, 101100, 101110, 101111, .... This factorization is illustrated in table D.4.

| factors | string |
|---|---|
| $p_1$ | 0 |
| $p_2$ | 1 |
| $p_1^2$ | 00 |
| $p_1 p_2$ | 01 |
| $p_2^2$ | 11 |
| $p_3$ | 10 |
| $p_1^3$ | 000 |
| $p_1^2 p_2$ | 001 |
| $p_1 p_2^2$ | 011 |
| $p_2^3$ | 111 |
| $p_1 p_3$ | 010 |
| $p_2 p_3$ | 110 |
| $p_4$ | 100 |
| $p_5$ | 101 |

| factors | string |
|---|---|
| $p_1^4$ | 0000 |
| $p_1^3 p_2$ | 0001 |
| $p_1^2 p_2^2$ | 0011 |
| $p_1 p_2^3$ | 0111 |
| $p_2^4$ | 1111 |
| $p_1^2 p_3$ | 0010 |
| $p_1 p_2 p_3$ | 0110 |
| $p_2^2 p_3$ | 1110 |
| $p_3^2$ | 1010 |
| $p_1 p_4$ | 0100 |
| $p_2 p_4$ | 1100 |
| $p_1 p_5$ | 0101 |
| $p_2 p_5$ | 1101 |
| $p_6$ | 1000 |
| $p_7$ | 1001 |
| $p_8$ | 1011 |

| factors | string |
|---|---|
| $p_1^5$ | 00000 |
| $p_1^4 p_2$ | 00001 |
| $p_1^3 p_2^2$ | 00011 |
| $p_1^2 p_2^3$ | 00111 |
| $p_1 p_2^4$ | 01111 |
| $p_2^5$ | 11111 |
| $p_1^3 p_3$ | 00010 |
| $p_1^2 p_2 p_3$ | 00110 |
| $p_1 p_2^2 p_3$ | 01110 |
| $p_2^3 p_3$ | 11110 |
| $p_1 p_3^2$ | 01010 |
| $p_2 p_3^2$ | 11010 |
| $p_1^2 p_4$ | 00100 |
| $p_1 p_2 p_4$ | 01100 |
| $p_2^2 p_4$ | 11100 |
| $p_3 p_4$ | 10100 |

| factors | string |
|---|---|
| $p_1^2 p_5$ | 00101 |
| $p_1 p_2 p_5$ | 01101 |
| $p_2^2 p_5$ | 11101 |
| $p_3 p_5$ | 10101 |
| $p_1 p_6$ | 01000 |
| $p_2 p_6$ | 11000 |
| $p_1 p_7$ | 01001 |
| $p_2 p_7$ | 11001 |
| $p_1 p_8$ | 01011 |
| $p_2 p_8$ | 11011 |
| $p_9$ | 10000 |
| $p_{10}$ | 10001 |
| $p_{11}$ | 10010 |
| $p_{12}$ | 10011 |
| $p_{13}$ | 10110 |
| $p_{14}$ | 10111 |

**Table D.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 $p_1 = 0$, $p_2 = 1$, $p_3 = 10$, $p_4 = 100$, ... , $p_{14} = 10111$.

### D.2.1 Prime factorization for spectral determinants

Following sect. D.2, the spectral determinant cycle expansions is obtained by expanding $F$ as a multinomial in prime cycle weights $t_p$

$$F = \prod_p \sum_{k=0}^{\infty} C_{p^k} t_p^k = \sum_{k_1 k_2 k_3 \cdots = 0}^{\infty} \tau_{p_1^{k_1} p_2^{k_2} p_3^{k_3} \cdots} \tag{D.12}$$

where the sum goes over all pseudocycles. In the above we have defined

$$\tau_{p_1^{k_1} p_2^{k_2} p_3^{k_3} \cdots} = \prod_{i=1}^{\infty} C_{p_i^{k_i}} t_{p_i}^{k_i} \ . \tag{D.13}$$

[exercise 17.10]

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. D.2, the relation of the cycle expansion (D.12) to the product formula (17.9) 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, $t_{pp\ldots p} = t_p^r$, 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 1-$d$, given by the Euler formula (21.34). In higher dimensions $C_{p^k}$ can be evaluated by expanding (17.9),

$F(z) = \prod_p F_p$, 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

$$\begin{aligned} F_p &= \sum_{r=1}^{\infty} C_k t^k, \quad C_k = (-)^k c_k / D_k, \\ D_k &= \prod_{r=1}^{k} d_r = \prod_{a=1}^{d} \prod_{r=1}^{k} (1 - u_a^r) \end{aligned} \tag{D.14}$$

where evaluation of $c_k$ requires a certain amount of not too luminous algebra:

$$\begin{aligned} c_0 &= 1 \\ c_1 &= 1 \\ c_2 &= \frac{1}{2}\left(\frac{d_2}{d_1} - d_1\right) = \frac{1}{2}\left(\prod_{a=1}^{d}(1+u_a) - \prod_{a=1}^{d}(1-u_a)\right) \\ c_3 &= \frac{1}{3!}\left(\frac{d_2 d_3}{d_1^2} + 2 d_1 d_2 - 3 d_3\right) \\ &= \frac{1}{6}\left(\prod_{a=1}^{d}(1 + 2u_a + 2u_a^2 + u_a^3)\right. \\ &\quad \left. + 2\prod_{a=1}^{d}(1 - u_a - u_a^2 + u_a^3) - 3\prod_{a=1}^{d}(1 - u_a^3)\right) \end{aligned}$$

*etc.*. For example, for a general 2-dimensional map we have

$$F_p = 1 - \frac{1}{D_1} t + \frac{u_1 + u_2}{D_2} t^2 - \frac{u_1 u_2 (1+u_1)(1+u_2) + u_1^3 + u_2^3}{D_3} t^3 + \dots . \tag{D.15}$$

We discuss the convergence of such cycle expansions in sect.I.4.

With $\tau_{p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n}}$ defined as above, the prime factorization of symbol strings is unique in the sense that *each symbol string can be written as a unique concatenation of prime strings*, up to a convention on ordering of primes. This factorization is a nontrivial example of the utility of generalized Möbius inversion, sect.D.2.

How is the factorization of sect. D.2 used in practice? Suppose we have computed (or perhaps even measured in an experiment) all prime cycles up to length $n$, i.e., we have a list of $t_p$'s and the corresponding fundamental 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

$$F = \sum_n \tau_n \tag{D.16}$$

where $n = s_1 s_2 \cdots s_j$, $s_i$ range over the alphabet, in the present case $\{0, 1\}$. Factorizing every string $n = s_1 s_2 \cdots s_j = p_1^{k_1} p_2^{k_2} \cdots p_j^{k_j}$ as in table D.4, and substituting $\tau_{p_1^{k_1} p_2^{k_2} \cdots}$ we obtain a multinomial approximation to $F$. For example, $\tau_{001001010101} = \tau_{\underline{001}\,\underline{001}\,\underline{01}\,\underline{01}\,\underline{01}} = \tau_{001^2}\tau_{01^3}$, and $\tau_{01^3}$, $\tau_{001^2}$ 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 (D.12) of the total cycle length $n$. Setting $C_{p^k} t_p^k = z^{n_p k}$ in (D.12), 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 (D.12) 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 length $n$ in the expansion (D.12).

# Appendix E

# Counting itineraries

## E.1 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 (18.7),

$$
\begin{aligned}
c_4 \;=\; & -t_{0001} - t_{0011} - t_{0111} - t_0 t_{01} t_1 \\
& + t_0 t_{001} + t_0 t_{011} + t_{001} t_1 + t_{011} t_1 \,. \qquad \text{(E.1)}
\end{aligned}
$$

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 E.2]

Such counting rules arise from the identity

$$
\prod_p \left(1 + t_p\right) = \prod_p \frac{1 - {t_p}^2}{1 - t_p} \,. \qquad \text{(E.2)}
$$

Substituting $t_p = z^{n_p}$ and using (13.15) we obtain for unrestricted symbol dynamics with $N$ letters

$$
\prod_p^{\infty} (1 + z^{n_p}) = \frac{1 - Nz^2}{1 - Nz} = 1 + Nz + \sum_{k=2}^{\infty} z^k \left(N^k - N^{k-1}\right)
$$

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 $(N-1)N^{k-1}$ (of which half carry a minus sign).

[exercise E.4]

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.

## Exercises

E.1. **Lefschetz zeta function.** Elucidate the relation between the topological zeta function and the Lefschetz zeta function.

E.2. **Counting the 3-disk pinball counterterms.** Verify that the number of terms in the 3-disk pinball curvature expansion (18.35) is given by

$$\begin{aligned}\prod_p \left(1+t_p\right) &= \frac{1-3z^4-2z^6}{1-3z^2-2z^3} = 1+3z^2+2z^3+\frac{z^4(6+12z+2z^2)}{1-3z^2-2z^3} \\ &= 1+3z^2+2z^3+6z^4+12z^5+20z^6+48z^7+84z^8+184z^9+ \qquad \text{(E.3)}\end{aligned}$$

This means that, for example, $c_6$ has a total of 20 terms, in agreement with the explicit 3-disk cycle expansion (18.36).

E.3. **Cycle expansion denominators****. Prove that the denominator of $c_k$ is indeed $D_k$, as asserted (D.14).

E.4. **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 (10.6) 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 (13.5) collapse to $tr(T^k) = (t_0+t_1)^k$, and $1/\zeta$ is simply

$$\prod_p \left(1-t_p\right) = 1-t_0-t_1 \qquad \text{(E.4)}$$

Substituting into the identity

$$\prod_p \left(1+t_p\right) = \prod_p \frac{1-{t_p}^2}{1-t_p}$$

we obtain

$$\begin{aligned}\prod_p \left(1+t_p\right) &= \frac{1-t_0^2-t_1^2}{1-t_0-t_1} = 1+t_0+t_1+\frac{2t_0t_1}{1-t_0-t_1} \\ &= 1+t_0+t_1+\sum_{n=2}^{\infty}\sum_{k=1}^{n-1} 2\binom{n-2}{k-1} t_0^k t_1^{n-k} \qquad \text{(E.5)}\end{aligned}$$

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 $M_{n,k}$ $(n = 1,2,\ldots;\ k = \{0,1\}$ for $n=1$; $k = 1,\ldots,n-1$ for $n \geq 2)$ the number of prime $n$-cycles whose labels contain $k$ zeros, use binomial string counting and Möbius inversion and obtain

$$\begin{aligned}M_{1,0} &= M_{1,1} = 1 \\ nM_{n,k} &= \sum_{m\left|\frac{n}{k}\right.} \mu(m)\binom{n/m}{k/m}, \qquad n \geq 2\,, k = 1,\ldots,n-\end{aligned}$$

where the sum is over all $m$ which divide both $n$ and $k$.

# Appendix F

# Finding cycles

(C. Chandre)

## F.1 Newton-Raphson method

### F.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

$$x' = g(x) = x - (J(x) - 1)^{-1} (f(x) - x) .$$

The linearization of $g$ near $x_*$ leads to

$$x_* + \epsilon' = x_* + \epsilon - (J(x_*) - \mathbf{1})^{-1} (f(x_*) + J(x_*)\epsilon - x_* - \epsilon) + O\left(\|\epsilon\|^2\right),$$

where $\epsilon = x - x_*$. Therefore,

$$x' - x_* = O\left((x - x_*)^2\right).$$

After $n$ steps and if the initial guess $x_0$ is close to $x_*$, the error decreases super-exponentially

$$g^n(x_0) - x_* = O\left((x_0 - x_*)^{2^n}\right).$$

### F.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

$$\begin{pmatrix} \mathbf{1} & & & -Df(x_n) \\ -Df(x_1) & \mathbf{1} & & \\ & \cdots & \mathbf{1} & \\ & & -Df(x_{n-1}) & \mathbf{1} \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \\ \cdots \\ \delta_n \end{pmatrix} = - \begin{pmatrix} F_1 \\ F_2 \\ \cdots \\ F_n \end{pmatrix}, \tag{F.1}$$

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 starightforward algebra, the vectors $\delta_m$ are expressed as functions of the vectors $F_m$:

$$\delta_m = -\sum_{k=1}^{m} \beta_{k,m-1} F_k - \beta_{1,m-1} \left(\mathbf{1} - \beta_{1,n}\right)^{-1} \left( \sum_{k=1}^{n} \beta_{k,n} F_k \right), \tag{F.2}$$

for $m = 1, \ldots, n$, 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 the inversing of a $[d \times d]$ matrix $\mathbf{1} - Df(x_n)Df(x_{n-1}) \cdots Df(x_1)$.

## F.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 notice that $x_*$ is also a fixed point of $g$. Consider the stability matrix at the fixed point $x_*$

$$A_g = \left. \frac{dg}{dx} \right|_{x=x_*} = 1 + \gamma C(A_f - 1).$$

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:

$$\tilde{A}_g = 1 + \gamma \tilde{C}(\tilde{A}_f - 1),$$

where $\tilde{A}_f$ is diagonal with elements $\mu_i$. 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

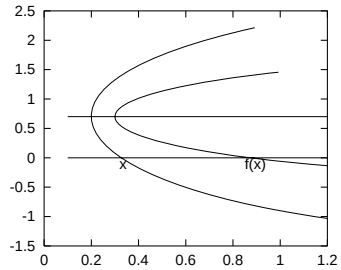



**Figure F.1:** Illustration of the optimal Poincaré surface. The original surface $y = 0$ yields a large distance $x - f(x)$ for the Newton iteration. A much better choice is $y = 0.7$.

$\gamma_i = 1 + \gamma\epsilon_i(\mu_i - 1)$. The choice of $\gamma$ and $\epsilon_i$ is such that $|\gamma_i| < 1$. It is easy to see that if $\mathrm{Re}(\mu_i) < 1$ one has to choose $\epsilon_i = 1$, and if $\mathrm{Re}(\mu_i) > 1$, $\epsilon_i = -1$. If $\lambda$ is chosen such that

$$0 < \gamma < \min_{i=1,\ldots,d} \frac{2|\mathrm{Re}(\mu_i) - 1|}{|\mu_i - 1|^2},$$

all the eigenvalues of $A_g$ have modulus less that one. The contraction rate at the fixed point for the map $g$ is then $\max_i |1 + \gamma\epsilon_i(\mu_i - 1)|$. We notice that if $\mathrm{Re}(\mu_i) = 1$, it is not possible to stabilize $x_*$ by the set of matrices $\gamma C$.
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 $\mathrm{Re}(\mu_1)\mathrm{Re}(\mu_2) \le \det Df < 1$. The case $(\mathrm{Re}(\mu_1) > 1, \mathrm{Re}(\mu_2) > 1)$ which is stabilized by $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ has to be discarded. The minimal set is reduced to three matrices.

### F.2.1 Newton method with optimal surface of section

(F. Christiansen)

In some systems it might be hard to find a good starting guess for a fixed point, something that could happen 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. In figure F.1 there is an illustration of this. 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 the 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 might in this context mean a Hamiltonian system with 3 degrees of freedom.) But clearly we could easily have 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 see by eye the best surface in

higher dimensional systems. The way to proceed is as follows: We want to have a minimal distance between our initial guess $x$ and the image of this $f(x)$. 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 we minimize $d$ 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 choose to define a Poincaré surface of section as the hyper-plane that goes through $x$ and is normal to the flow at $x$. In other words the surface of section is determined by

$$(x' - x) \cdot v(x) = 0. \tag{F.3}$$

Note that $f(x)$ lies on this surface. This surface of section is optimal in the sense that a close return on the surface is a local minimum of the distance between $x$ and $f^t(x)$. But more importantly, the part of the stability matrix that describes linearization perpendicular to the flow is exactly the stability of the flow in the surface of section when $f(x)$ is close to $x$. In this method, the Poincaré surface changes with each iteration of the Newton scheme. Should we later want to put the fixed point on a specific Poincaré surface it will only be a matter of moving along the trajectory.

# Appendix G

# Transport of vector fields

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.

## G.1 Evolution operator for Lyapunov exponents

Lyapunov exponents were introduced and computed for 1-$d$ maps in sect.15.3.2. For higher-dimensional flows only the fundamental 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 (G.4) whose spectral determinant (G.8) yields the leading Lyapunov exponent of a $d$-dimensional flow (and is entire for Axiom A flows).

The key idea is to extending 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 $\eta(0) \in \mathbf{T}\mathcal{M}_x$, and let the flow transport it along the trajectory $x(t) = f^t(x_0)$.

The dynamics in the $(x, \eta) \in U \times TU_x$ space is governed by the system of equations of variations [1]:

$$\dot{x} = \mathbf{v}(x)\,, \quad \dot{\eta} = \mathbf{D}\mathbf{v}(x)\eta\,.$$

Here $\mathbf{Dv}(x)$ is the derivative matrix of the flow. We write the solution as

$$x(t) = f^t(x_0)\,, \quad \eta(t) = M^t(x_0)\cdot\eta_0\,, \tag{G.1}$$

with the tangent space vector $\eta$ transported by the stability matrix $M^t(x_0) = \partial x(t)/\partial x_0$.

As explained in sect. 4.1, the growth rate of this vector is multiplicative along the trajectory and can be represented as $\eta(t) = |\eta(t)|/|\eta(0)|\mathbf{u}(t)$ where $\mathbf{u}(t)$ is a "unit" vector in some norm $||.||$. 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 stability eigenvalues by adjoining the $d$-dimensional transverse tangent space $\eta \in \mathbf{T}\mathcal{M}_x;\ \eta(x)\mathbf{v}(x) = 0$ 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 $\eta$ we introduce a homogeneous differentiable scalar function $g(\eta) = ||\eta||$. It has the property $g(\Lambda\eta) = |\Lambda|g(\eta)$ for any $\Lambda$. An example is the projection of a vector to its $d$th component

$$g\begin{pmatrix}\eta_1\\ \eta_2\\ \cdots\\ \eta_d\end{pmatrix} = |\eta_d|\,.$$

Any vector $\eta \in TU_x$ can now be represented by the product $\eta = \Lambda\mathbf{u}$, where $\mathbf{u}$ is a "unit" vector in the sense that its norm is $||\mathbf{u}|| = 1$, and the factor

$$\Lambda^t(x_0, \mathbf{u}_0) = g(\eta(t)) = g(M^t(x_0)\cdot\mathbf{u}_0) \tag{G.2}$$

is the multiplicative "stretching" factor.

Unlike the leading eigenvalue of the Jacobian the stretching factor is multiplicative along the trajectory:

$$\Lambda^{t'+t}(x_0, \mathbf{u}_0) = \Lambda^{t'}(x(t), \mathbf{u}(t))\,\Lambda^t(x_0, \mathbf{u}_0).$$

[exercise G.1]

The $\mathbf{u}$ evolution constrained to $E\mathbf{T}_{g,x}$, the space of unit transverse tangent vectors, is given by rescaling of (G.1):

$$\mathbf{u}' = R^t(x, \mathbf{u}) = \frac{1}{\Lambda^t(x, \mathbf{u})} M^t(x)\cdot\mathbf{u}\,. \tag{G.3}$$

Eqs. (G.1), (G.2) and (G.3) enable us to define a *multiplicative* evolution operator on the extended space $U \times E\mathbf{T}_{g,x}$

$$\mathcal{L}^t(x', \mathbf{u}'; x, \mathbf{u}) = \delta\Big(x' - f^t(x)\Big)\,\frac{\delta(\mathbf{u}' - R^t(x, \mathbf{u}))}{|\Lambda^t(x, \mathbf{u})|^{\beta-1}}\,, \tag{G.4}$$

where $\beta$ is a variable.

To evaluate the expectation value of $\log|\Lambda^t(x,\mathbf{u})|$ which is the Lyapunov exponent we again have to take the proper derivative of the leading eigenvalue of (G.4). In order to derive the trace formula for the operator (G.4) we need to evaluate $\mathrm{Tr}\,\mathcal{L}^t = \int dx d\mathbf{u}\, \mathcal{L}^t(\mathbf{u},x;\mathbf{u},x)$. The $\int dx$ integral yields a weighted sum over prime periodic orbits $p$ and their repetitions $r$:

$$
\begin{aligned}
\mathrm{Tr}\,\mathcal{L}^t &= \sum_p T_p \sum_{r=1}^{\infty} \frac{\delta\Big(t - rT_p\Big)}{|\det(1-M_p^r)|}\Delta_{p,r}, \\
\Delta_{p,r} &= \int_g d\mathbf{u}\, \frac{\delta\Big(\mathbf{u} - R^{T_p r}(x_p,\mathbf{u})\Big)}{|\Lambda^{T_p r}(x_p,\mathbf{u})|^{\beta-1}}\,,
\end{aligned}
\tag{G.5}
$$

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 cycle 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

$$
\mathbf{e}_i(x(T+t)) = \mathbf{e}_i(x(t))\,, \quad i = 1,...,d,
$$

corresponding to the $d$ unit eigenvectors $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_d\}$ of the transverse stability matrix, with "stretching" factors (G.2) given by its eigenvalues

$$
M_p(x)\cdot\mathbf{e}_i(x) = \Lambda_{p,i}\,\mathbf{e}_i(x)\,, \quad i = 1,...,d. \qquad \text{(no summation on } i\text{)}
$$

The $\int d\mathbf{u}$ integral in (G.5) picks up contributions from these periodic solutions. In order to compute the stability of the $i$th eigendirection solution, it is convenient to expand the variation around the eigenvector $\mathbf{e}_i$ in the stability matrix eigenbasis $\delta\mathbf{u} = \sum \delta u_\ell\,\mathbf{e}_\ell$ . The variation of the map (G.3) at a complete period $t = T$ is then given by

$$
\begin{aligned}
\delta R^T(\mathbf{e}_i) &= \frac{M\cdot\delta\mathbf{u}}{g(M\cdot\mathbf{e}_i)} - \frac{M\cdot\mathbf{e}_i}{g(M\cdot\mathbf{e}_i)^2}\left(\frac{\partial g(\mathbf{e}_i)}{\partial\mathbf{u}}\cdot M\cdot\delta\mathbf{u}\right) \\
&= \sum_{k\neq i}\frac{\Lambda_{p,k}}{\Lambda_{p,i}}\left(\mathbf{e}_k - \mathbf{e}_i\frac{\partial g(\mathbf{e}_i)}{\partial u_k}\right)\delta u_k\,.
\end{aligned}
\tag{G.6}
$$

The $\delta u_i$ 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 $\delta\mathbf{u}$ must satisfy

$$
g(\mathbf{u}+\delta\mathbf{u}) = g(\mathbf{u}) = 1 \quad\Longrightarrow\quad \sum_{\ell=1}^{d}\delta u_\ell\frac{\partial g(\mathbf{u})}{\partial u_\ell} = 0\,,
$$

so the allowed variations are of form

$$\delta\mathbf{u} = \sum_{k\neq i}\left(\mathbf{e}_k - \mathbf{e}_i\frac{\partial g(\mathbf{e}_i)}{\partial u_k}\right)c_k\,, \quad |c_k| \ll 1\,,$$

and in the neighborhood of the $\mathbf{e}_i$ eigenvector the $\int d\mathbf{u}$ integral can be expressed as

$$\int_g d\mathbf{u} = \int \prod_{k\neq i} dc_k\,.$$

Inserting these variations into the $\int d\mathbf{u}$ integral we obtain

$$\begin{aligned}\int_g d\mathbf{u} \qquad & \delta\Big(\mathbf{e}_i + \delta\mathbf{u} - R^T(\mathbf{e}_i) - \delta R^T(\mathbf{e}_i) + \ldots\Big)\\ = \qquad & \int \prod_{k\neq i} dc_k\,\delta((1-\Lambda_k/\Lambda_i)c_k + \ldots)\\ & = \prod_{k\neq i}\frac{1}{|1-\Lambda_k/\Lambda_i|}\,,\end{aligned}$$

and the $\int d\mathbf{u}$ trace (G.5) becomes

$$\Delta_{p,r} = \sum_{i=1}^{d}\frac{1}{|\,\Lambda_{p,i}^r\,|^{\beta-1}}\prod_{k\neq i}\frac{1}{|\,1-\Lambda_{p,k}^r/\Lambda_{p,i}^r\,|}\,. \tag{G.7}$$

The corresponding spectral determinant is obtained by observing that the Laplace transform of the trace (16.23) is a logarithmic derivative $\mathrm{Tr}\,\mathcal{L}(s) = -\frac{d}{ds}\log F(s)$ of the spectral determinant:

$$F(\beta,s) = \exp\left(-\sum_{p,r}\frac{e^{sT_pr}}{r\,|\det(1-M_p^r)\,|}\Delta_{p,r}(\beta)\right). \tag{G.8}$$

This determinant is the central result of this section. Its zeros correspond to the eigenvalues of the evolution operator (G.4), and can be evaluated by the cycle expansion methods.

The leading zero of (G.8) is called "pressure" (or free energy)

$$P(\beta) = s_0(\beta). \tag{G.9}$$

The average Lyapunov exponent is then given by the first derivative of the pressure at $\beta = 1$:

$$\overline{\lambda} = P'(1). \tag{G.10}$$

The simplest application of (G.8) is to 2-dimensional hyperbolic Hamiltonian maps. The stability eigenvalues are related by $\Lambda_1 = 1/\Lambda_2 = \Lambda$, and the spectral determinant is given by

$$\begin{aligned} F(\beta, z) &= \exp\left(-\sum_{p,r} \frac{z^{rn_p}}{r \mid \Lambda_p^r \mid} \frac{1}{(1-1/\Lambda_p^r)^2} \Delta_{p,r}(\beta)\right) \\ \Delta_{p,r}(\beta) &= \frac{\mid \Lambda_p^r \mid^{1-\beta}}{1-1/\Lambda_p^{2r}} + \frac{\mid \Lambda_p^r \mid^{\beta-3}}{1-1/\Lambda_p^{2r}} \,. \end{aligned} \tag{G.11}$$

The dynamics (G.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 (G.7) is also entire.

In case of maps it is practical to introduce the logarithm of the leading zero and to call it "pressure"

$$P(\beta) = \log z_0(\beta). \tag{G.12}$$

The average of the Lyapunov exponent of the map is then given by the first derivative of the pressure at $\beta = 1$:

$$\overline{\lambda} = P'(1). \tag{G.13}$$

By factorizing the determinant (G.11) into products of zeta functions we can conclude that the leading zero of the (G.4) can also be recovered from the leading zeta function

$$1/\zeta_0(\beta, z) = \exp\left(-\sum_{p,r} \frac{z^{rn_p}}{r|\Lambda_p^r|^\beta}\right). \tag{G.14}$$

This zeta function plays a key role in thermodynamic applications as we will will see in Chapter 22.

## G.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

$$\partial_t \mathbf{V} + \mathbf{u} \cdot \nabla \mathbf{V} - \mathbf{V} \cdot \nabla \mathbf{u} = 0, \tag{G.15}$$

where $\mathbf{u}(x,t)$ 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

$$\mathbf{V}(x,t) \sim \mathbf{V}(x)e^{\nu t}, \tag{G.16}$$

where $\nu$ 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 (G.15) 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

$$\mathbf{V}(\mathbf{x},t) = \mathbf{J}(\mathbf{a},t)\mathbf{V}(\mathbf{a},0) \quad , \tag{G.17}$$

where $\mathbf{J}(\mathbf{a},t) = \partial(\mathbf{x})/\partial(\mathbf{a})$ is the fundamental 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 (G.17)

$$V_i(\mathbf{x},t) = \sum_j \int d^3\mathbf{a}\ \mathcal{L}^t_{ij}(\mathbf{x},\mathbf{a})V_j(\mathbf{a},0) \quad , \tag{G.18}$$

with

$$\mathcal{L}^t_{ij}(\mathbf{x},\mathbf{a}) = \delta(\mathbf{a} - f^{-t}(\mathbf{x}))\frac{\partial x_i}{\partial a_j} \quad . \tag{G.19}$$

For large times, the effect of $\mathcal{L}^t$ is dominated by its leading eigenvalue, $e^{\nu_0 t}$ with $Re(\nu_0) > Re(\nu_i)$, $i = 1,2,3,\ldots$. 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

$$\mathbf{Tr}\mathcal{L}^t = \sum_p T_p \sum_{r=1}^{\infty} \frac{\mathrm{tr}\, M_p^r}{\left|\det\left(\mathbf{1} - M_p^{-r}\right)\right|}\delta(t - rT_p). \tag{G.20}$$

We can construct the corresponding spectral determinant as usual

$$F(s) = \exp\left[-\sum_p \sum_{r=1}^{\infty} \frac{1}{r}\frac{\mathrm{tr}\, M_p^r}{\left|\det\left(\mathbf{1} - M_p^{-r}\right)\right|}e^{srT_p}\right] \quad . \tag{G.21}$$

Note that in this formuli we have omitted a term arising from the Jacobian transformation along the orbit which would give $1 + \mathrm{tr}\ M_p^r$ 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 (G.19).

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 $\Lambda_p$ (with $|\Lambda_p| > 1$), and one contracting eigenvalue $\lambda_p$ (with $|\lambda_p| < 1$) the determinant may be expanded as follows:

$$\left|\det\left(\mathbf{1} - M_p^{-r}\right)\right|^{-1} = |(1-\Lambda_p^{-r})(1-\lambda_p^{-r})|^{-1} = |\lambda_p|^r \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \Lambda_p^{-jr}\lambda_p^{kr} \quad . \tag{G.22}$$

With this decomposition we can rewrite the exponent in (G.21) as

$$\sum_p \sum_{r=1}^{\infty} \frac{1}{r}\frac{(\lambda_p^r + \Lambda_p^r)e^{srT_p}}{\left|\det\left(\mathbf{1} - M_p^{-r}\right)\right|} = \sum_p \sum_{j,k=0}^{\infty}\sum_{r=1}^{\infty} \frac{1}{r}\left(|\lambda_p|\Lambda_p^{-j}\lambda_p^k e^{sT_p}\right)^r (\lambda_p^r + \Lambda_p^r) \quad , \tag{G.23}$$

which has the form of the expansion of a logarithm:

$$\sum_p \sum_{j,k} \left[\log\left(1 - e^{sT_p}|\lambda_p|\Lambda_p^{1-j}\lambda_p^k\right) + \log\left(1 - e^{sT_p}|\lambda_p|\Lambda_p^{-j}\lambda_p^{1+k}\right)\right] \quad . \tag{G.24}$$

The spectral determinant is therefore of the form,

$$F(s) = F_e(s)F_c(s) \quad , \tag{G.25}$$

where

$$F_e(s) = \prod_p \prod_{j,k=0}^{\infty}\left(1 - t_p^{(jk)}\Lambda_p\right) \quad , \tag{G.26}$$

$$F_c(s) = \prod_p \prod_{j,k=0}^{\infty}\left(1 - t_p^{(jk)}\lambda_p\right), \tag{G.27}$$

with

$$t_p^{(jk)} = e^{sT_p}|\lambda_p|\frac{\lambda_p^k}{\Lambda_p^j} \quad . \tag{G.28}$$

The two factors present in $F(s)$ correspond to the expanding and contracting exponents. (Had we not neglected a term in (G.21), there would be a third factor corresponding to the translation.)

For 2-$d$ Hamiltonian volume preserving systems, $\lambda = 1/\Lambda$ and (G.26) reduces to

$$F_e(s) = \prod_p \prod_{k=0}^{\infty} \left(1 - \frac{t_p}{\Lambda_p^{k-1}}\right)^{k+1} , \quad t_p = \frac{e^{sT_p}}{|\Lambda_p|} \quad . \tag{G.29}$$

With $\sigma_p = \Lambda_p/|\Lambda_p|$, the Hamiltonian zeta function (the $j = k = 0$ part of the product (G.27)) is given by

$$1/\zeta_{dyn}(s) = \prod_p \left(1 - \sigma_p e^{sT_p}\right) . \tag{G.30}$$

This is a curious formula — the zeta function depends only on the return times, not on the eigenvalues of the cycles. Furthermore, the identity,

$$\frac{\Lambda + 1/\Lambda}{|(1-\Lambda)(1-1/\Lambda)|} = \sigma + \frac{2}{|(1-\Lambda)(1-1/\Lambda)|} \quad ,$$

when substituted into (G.25), leads to a relation between the vector and scalar advection spectral determinants:

$$F_{dyn}(s) = F_0^2(s)/\zeta_{dyn}(s) . \tag{G.31}$$

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 (G.29) and (G.30) for the dynamo determinants by simply making the substitution

$$z^{n_p} = e^{sT_p} \quad , \tag{G.32}$$

where $n_p$ is the integer order of the cycle. Then we find the spectral determinant $F_e(z)$ given by equation (G.29) but with

$$t_p = \frac{z^{n_p}}{|\Lambda_p|} \tag{G.33}$$

for the weights, and

$$1/\zeta_{dyn}(z) = \Pi_p \left(1 - \sigma_p z^{n_p}\right) \tag{G.34}$$

for the zeta-function

For *maps* with finite Markov partition the inverse zeta function (G.34) 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

$$1/\zeta_{dyn}(s) = 1. \tag{G.35}$$

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, (G.31) implies that the dynamo zeta function is a ratio of two entire determinants:

$$1/\zeta_{dyn}(s) = F_{dyn}(s)/F_0^2(s)\,. \tag{G.36}$$

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 G.2]

## Commentary

**Remark G.1** Dynamo zeta. The dynamo zeta (G.34) has been introduced by Aurell and Gilbert [2] and reviewed in ref. [3]. Our exposition follows ref. [19].

## Exercises

G.1. **Stretching factor.** Prove the multiplicative property of the stretching factor (G.2). Why should we extend the phase space with the tangent space?

G.2. **Dynamo rate.** Suppose that the fluid dynamics is highly dissipative and can be well approximated by the piecewise linear map

$$f(x) = \begin{cases} 1 + ax & \text{if} \quad x < 0, \\ 1 - bx & \text{if} \quad x > 0, \end{cases} \tag{G.37}$$

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

$$1/\zeta_{dyn}(s) = 1 - e^{sT_a} + e^{sT_b}. \tag{G.38}$$

Show also that the escape rate is the leading zero of

$$1/\zeta_0(s) = 1 - e^{sT_a}/a - e^{sT_b}/b. \tag{G.39}$$

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?

## References


[G.1] Ya.B. Pesin, *Uspekhi Mat. Nauk* **32,** 55 (1977), [*Russian Math. Surveys* **32,** 55 (1977)].

[G.2] E. Aurell and A. Gilbert, *Geophys. & Astrophys. Fluid Dynamics* (1993).

[G.3] S. Childress and A.D. Gilbert *Stretch, Twist, Fold: The Fast Dynamo*, *Lecture Notes in Physics* **37** (Springer Verlag, Berlin 1995).

[G.4] J. Balatoni and A. Renyi, *Publi. Math. Inst. Hung. Acad. Sci.* **1**, 9 (1956); english translation **1**, 588 (Akademia Budapest, 1976).

[G.5] R. Benzi, G. Paladin, G. Parisi and A. Vulpiani, *J. Phys.* **A17**, 3521 (1984).

[G.6] Even though the thermodynamic formalism is of older vintage (we refer the reader to ref. [28] for a comprehensive, and incomprehensible overview), we adhere here to the notational conventions of ref. [7], more frequently employed in the physics literature.

[G.7] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia and B.I. Shraiman, *Phys. Rev.* **A107**, 1141 (1986).

[G.8] P. Grassberger, *Phys. Lett.* **97A**, 227 (1983); **107A**, 101 (1985); H.G.E. Hentschel and I. Procaccia, *Physica* **8D**, 435 (1983); R. Benzi, G. Paladin, G. Parisi and A. Vulpiani, em J. Phys. **A17**, 3521 (1984).

[G.9] P. Grassberger and I. Procaccia, *Phys. Rev. A* **31**, 1872 (1985).

[G.10] C. Shannon, *Bell System Technical Journal*, **27**, 379 (1948).

[G.11] H. Fujisaka, *Progr. Theor. Phys* **70**, 1264 (1983).

[G.12] M. Barnsley, *Fractals Everywhere*, (Academic Press, New York 1988).

[G.13] A.S. Pikovsky, unpublished.

[G.14] C. Beck, "Higher correlation functions of chaotic dynamical systems - a graph theoretical approach," *Nonlinearity* **4**, 1131 (1991); to be published.

[G.15] The 4-point correlation function is given in ref. [14].

[G.16] G. Hackenbroich and F. von Oppen, "Semiclassical theory of transport in antidot lattices," *Z. Phys.* **B 97**, 157 (1995).

[G.17] M.J. Feigenbaum, *J. Stat. Phys.* **21**, 669 (1979); reprinted in ref. [5].

[G.18] P. Szépfalusy, T. Tél, A. Csordás and Z. Kovács, *Phys. Rev.* **A 36**, 3525 (1987).

# Appendix H

# Discrete symmetries of dynamics

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

$$\mathbf{M} = \lambda_1 \mathbf{P}_1 + \lambda_2 \mathbf{P}_2 + \cdots + \lambda_r \mathbf{P}_r \,,$$

which associates with each distinct root $\lambda_i$ of invariant matrix $\mathbf{M}$ a projection operator (H.17):

$$\mathbf{P}_i = \prod_{j \neq i} \frac{\mathbf{M} - \lambda_j \mathbf{1}}{\lambda_i - \lambda_j} \,.$$

Sects. H.3 and H.4 develop Fourier analysis as an application of the general theory of invariance groups and their representations.

## H.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 $g_1, g_2, g_3, \ldots$ 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

$$G = \{e, g_2, \ldots, g_{|G|}\},$$

where $|G|$, the number of elements, is the *order* of the group.

**Example H.1** ***Finite groups:*** *Some finite groups that frequently arise in applications:*

- $C_n$ *(also denoted $Z_n$): 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 $n!$.*

**Example H.2** ***Lie groups:*** *Some compact continuous groups that arise in dynamical systems applications:*

- $S^1$ *(also denoted $T^1$):* circle group *of dimension* $1$.
- $T_m = S^1 \times S^1 \cdots \times S^1$*: $m$-*torus*, of dimension $m$.*
- $SO(2)$*: rotations in the plane, dimension* $1$*. Isomorphic to* $S^1$.
- $O(2) = SO(2) \times D_1$*: group of rotations and reflections in the plane, of dimension* $1$.
- $U(1)$*: group of phase rotations in the complex plane, of dimension* $1$*. Isomorphic to* $SO(2)$.
- $SO(3)$*:* rotation group *of dimension* $3$.
- $SU(2)$*:* unitary group *of dimension* $3$*. Isomorphic to* $SO(3)$.
- $GL(n)$*:* general linear group *of invertible matrix transformations, dimension* $n^2$.
- $SO(n)$*:* special orthogonal group *of dimension* $n(n-1)/2$.
- $O(n) = SO(n) \times D_1$*:* orthogonal group *of dimension* $n(n-1)/2$.
- $Sp(n)$*:* symplectic group *of dimension* $n(n+1)/2$.
- $SU(n)$*:* special unitary group *of dimension* $n^2 - 1$.

**Example H.3** ***Cyclic and dihedral groups:*** *The cyclic group $C_n \subset SO(2)$ of order $n$ is generated by one element. For example, this element can be rotation through $2\pi/n$.*

*The dihedral group $D_n \subset O(2)$, $n > 2$, can be generated by two elements one at least of which must reverse orientation. For example, take $\sigma$ corresponding to reflection in the $x$-axis. $\sigma^2 = e$; such operation $\sigma$ is called an* involution. *$C$ to rotation through $2\pi/n$, then $D_n = \langle \sigma, C \rangle$, and the defining relations are $\sigma^2 = C^n = e$, $(C\sigma)^2 = e$.*

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 B.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

$$G_a{}^b x_b \equiv \sum_{b=1}^{n} G_a{}^b x_b \,, \tag{H.1}$$

unless explicitly stated otherwise.

**General linear transformations.** Let $GL(n, \mathbb{F})$ be the group of general linear transformations,

$$GL(n, \mathbb{F}) = \{\mathbf{g} : \mathbb{F}^n \to \mathbb{F}^n \,|\, \det(\mathbf{g}) \neq 0\} \,. \tag{H.2}$$

Under $GL(n, \mathbb{F})$ a basis set of $V$ is mapped into another basis set by multiplication with a $[n \times n]$ matrix $\mathbf{g}$ with entries in field $\mathbb{F}$ ($\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$),

$$\mathbf{e}'^{\,a} = \mathbf{e}^b (\mathbf{g}^{-1})_b{}^a \,.$$

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 $\mathbf{g}$ as a *standard rep* of $GL(n, \mathbb{F})$, and the space of all $n$-tuples $(x_1, x_2, \ldots, x_n)^T$, $x_i \in \mathbb{F}$ on which these matrices act as the *standard representation space* $V$.

Under a general linear transformation $\mathbf{g} \in GL(n, \mathbb{F})$, the row of basis vectors transforms by right multiplication as $\mathbf{e}' = \mathbf{e}\,\mathbf{g}^{-1}$, 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 $\mathbf{e}^T$ transforms as $\mathbf{e}'^T = \mathbf{g}^\dagger \mathbf{e}^T$, where the *dual rep* $\mathbf{g}^\dagger = (\mathbf{g}^{-1})^T$ is the transpose of the inverse of $\mathbf{g}$. This observation motivates introduction of a *dual* representation space $\overline{V}$, the space on which $GL(n, \mathbb{F})$ acts via the dual rep $\mathbf{g}^\dagger$.

**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 $\{\mathbf{e}^{(1)}, \cdots, \mathbf{e}^{(d)}\}$ is a (right) basis of $V$, then $\bar{V}$ is spanned by the *dual basis* (left basis) $\{\mathbf{e}_{(1)}, \cdots, \mathbf{e}_{(d)}\}$, the set of $n$ linear forms $\mathbf{e}_{(j)}$ such that

$$\mathbf{e}_{(i)} \cdot \mathbf{e}^{(j)} = \delta_i^j \,,$$

where $\delta_a^b$ 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

$$(y^1, y^2, \ldots, y^n) \,. \tag{H.3}$$

They transform under $GL(n, \mathbb{F})$ as

$$y'^a = (\mathbf{g}^\dagger)_b{}^a y^b \,. \tag{H.4}$$

For $GL(n, \mathbb{F})$ no complex conjugation is implied by the $^\dagger$ notation; that interpretation applies only to unitary subgroups of $GL(n, \mathbb{C})$. $\mathbf{g}$ can be distinguished from $\mathbf{g}^\dagger$ by meticulously keeping track of the relative ordering of the indices,

$$g_a^b \to g_a{}^b \,, \qquad (\mathbf{g}^\dagger)_a^b \to g^b{}_a \,. \tag{H.5}$$

**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 (H.3), so the components of defining space vectors, resp. dual vectors, are distinguished by lower, resp. upper indices:

$$\begin{aligned} x &= (x_1, x_2, \ldots, x_n) \,, & \mathbf{x} \in V \\ \bar{x} &= (x^1, x^2, \ldots, x^n) \,, & \bar{\mathbf{x}} \in \bar{V} \,. \end{aligned} \tag{H.6}$$

**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 $\mathbf{g}$

$$x'_a = g_a{}^b x_b \qquad a, b = 1, 2, \ldots, n \,. \tag{H.7}$$

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 $\mathbf{g}^\dagger$:

$$x'^a = x^b (\mathbf{g}^\dagger)_b{}^a = g^a{}_b x^b \,. \tag{H.8}$$

In the applications considered here, the group $G$ will almost always be assumed to be a subgroup of the *unitary group*, in which case $\mathbf{g}^{-1} = \mathbf{g}^\dagger$, and $^\dagger$ indicates hermitian conjugation:

$$(\mathbf{g}^\dagger)_a{}^b = (g_b{}^a)^* = g^b{}_a \,. \tag{H.9}$$

**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

$$(\mathbf{M}^\dagger)^a_b = M^a_b \,. \tag{H.10}$$

For a hermitian matrix there is no need to keep track of the relative ordering of indices, as $M_b{}^a = (\mathbf{M}^\dagger)_b{}^a = M^a{}_b$.

**Invariant vectors.** The vector $q \in V$ is an *invariant vector* if for any transformation $g \in G$

$$q = \mathbf{g}q \,. \tag{H.11}$$

If a bilinear form $\mathbf{M}(\bar{x}, y) = x^a M_a{}^b y_b$ is invariant for all $g \in G$, the matrix

$$M_a{}^b = g_a{}^c g^b{}_d M_c{}^d \tag{H.12}$$

is an *invariant matrix*. Multiplying with $g_b{}^e$ and using the unitary condition (H.9), we find that the invariant matrices *commute* with all transformations $g \in G$:

$$[\mathbf{g}, \mathbf{M}] = 0 \,. \tag{H.13}$$

**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

$$H(x, y, \bar{z}, \bar{r}, \bar{s}) = h^{ab}{}_{cde} x_b y_a s^e r^d z^c \,, \tag{H.14}$$

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 group on vector space.** We will now apply these abstract group definitions to the set of $[d \times d]$-dimensional non-singular matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \ldots \in GL(d)$ acting in a $d$-dimensional vector space $V \in \mathbb{R}^d$. The product of matrices $\mathbf{A}$ and $\mathbf{B}$ gives the matrix $\mathbf{C}$,

$$\mathbf{C}x = \mathbf{B}(\mathbf{A}x) = (\mathbf{BA})x \in V, \qquad \forall x \in V.$$

The identity of the group is the unit matrix $\mathbf{1}$ which leaves all vectors in $V$ unchanged. Every matrix in the group has a unique inverse.

**Matrix representation of a group.** Let us now map the abstract group $G$ *homeomorphically* on a group of matrices $\mathsf{D}(G)$ 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 $\mathbf{D}(g) \in \mathsf{D}(G)$.
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 is preserved: $\mathbf{D}(g_3 \circ (g_2 \circ g_1)) = \mathbf{D}(g_3)(\mathbf{D}(g_2)\mathbf{D}(g_1)) = (\mathbf{D}(g_3)(\mathbf{D}(g_2))\mathbf{D}(g_1)$.
4. The identity element $e \in G$ is mapped onto the unit matrix $\mathsf{D}(e) = \mathbf{1}$ 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 $\mathsf{D}(G)$ 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 $\mathbf{D}(g) \in \mathsf{D}(G)$ as $\mathbf{g}$, i.e., $x' = \mathbf{g}x$ corresponds to the matrix operation $x'_i = \sum_{j=1}^{d} \mathbf{D}(g)_{ij} x_j$.

**Character of a representation.** The character of $\chi_\alpha(g)$ of a $d$-dimensional representation $\mathbf{D}(g)$ of the group element $g \in G$ is defined as trace

$$\chi_\alpha(g) = \operatorname{tr} \mathbf{D}(g) = \sum_{i=1}^{d} \mathbf{D}_{ii}(g)\,.$$

Note that $\chi(e) = d$, since $\mathbf{D}_{ij}(e) = \delta_{ij}$ for $1 \le i, j \le d$.

**Faithful representations, factor group.** If the mapping $G$ on $\mathsf{D}(G)$ is an isomorphism, the representation is said to be *faithful*. In this case the order of the group of matrices $D(G)$ 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 $\mathsf{D}'(G)$ equivalent to the representation $\mathsf{D}(G)$ of the group $G$. The equivalent representations have the same structure, although the matrices look different. Because of the cylic nature of the trace the character of equivalent representations is the same

$$\chi(g) = \sum_{i=1}^{n} \mathbf{D}'_{ii}(g) = \mathrm{tr}\,\mathbf{D}'(g) = \mathrm{tr}\left(\mathbf{C}\mathbf{D}(g)\mathbf{C}^{-1}\right)\,.$$

**Regular representation of a finite group.** The *regular* representation of a group is a special representation that is defined as follows: Combine the elements of a finite group into a vector $\{g_1, g_2, \ldots, g_{|G|}\}$. Multiplication by any element $g_\nu$ permutes $\{g_1, g_2, \ldots, g_{|G|}\}$ entries. We can represent the element $g_\nu$ by the permutation it induces on the components of vector $\{g_1, g_2, \ldots, g_{|G|}\}$. Thus for $i, j = 1, \ldots, |G|$, we define the *regular representation*

$$\mathbf{D}_{ij}(g_\nu) = \begin{cases} \delta_{jl_i} & \text{if } g_\nu g_i = g_{l_i} \text{ with } l_i = 1, \ldots, |G|\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

In the regular representation the diagonal elements of all matrices are zero except for the identity element $g_\nu = e$ with $g_\nu g_i = g_i$. So in the regular representation the character is given by

$$\chi(g) = \begin{cases} |G| & \text{for} \quad g = e\,, \\ 0 & \text{for} \quad g \neq e\,. \end{cases}$$

## H.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.

### H.2.1 Projection operators

For $\mathbf{M}$ a hermitian matrix, there exists a diagonalizing unitary matrix $\mathbf{C}$ such that

$$\mathbf{CMC}^\dagger = \left(\begin{array}{c|c|c} \begin{matrix}\lambda_1 & \dots & 0\\ & \ddots & \\ 0 & \dots & \lambda_1\end{matrix} & 0 & 0 \\ \hline 0 & \begin{matrix}\lambda_2 & 0 & \dots & 0\\ 0 & \lambda_2 & & \\ \vdots & & \ddots & \vdots \\ 0 & & \dots & \lambda_2\end{matrix} & 0 \\ \hline 0 & 0 & \begin{matrix}\lambda_3 & \dots \\ \vdots & \ddots\end{matrix} \end{array}\right) . \tag{H.15}$$

Here $\lambda_i \neq \lambda_j$ are the $r$ distinct roots of the minimal *characteristic* (or *secular*) polynomial

$$\prod_{i=1}^{r}(\mathbf{M} - \lambda_i \mathbf{1}) = 0 . \tag{H.16}$$

In the matrix $\mathbf{C}(\mathbf{M} - \lambda_2 \mathbf{1})\mathbf{C}^\dagger$ the eigenvalues corresponding to $\lambda_2$ are replaced by zeroes:

$$\left(\begin{array}{c|c|c} \begin{matrix}\lambda_1 - \lambda_2 & \\ & \lambda_1 - \lambda_2\end{matrix} & & \\ \hline & \begin{matrix}0 & & \\ & \ddots & \\ & & 0\end{matrix} & \\ \hline & & \begin{matrix}\lambda_3 - \lambda_2 & & \\ & \lambda_3 - \lambda_2 & \\ & & \ddots\end{matrix} \end{array}\right) ,$$

and so on, so the product over all factors $(\mathbf{M} - \lambda_2\mathbf{1})(\mathbf{M} - \lambda_3\mathbf{1})\ldots$, with exception of the $(\mathbf{M} - \lambda_1\mathbf{1})$ factor, has nonzero entries only in the subspace associated with $\lambda_1$:

$$\mathbf{C}\prod_{j\neq 1}(\mathbf{M} - \lambda_j\mathbf{1})\mathbf{C}^\dagger = \prod_{j\neq 1}(\lambda_1 - \lambda_j)\left(\begin{array}{ccc|cccc} 1 & 0 & 0 & & & 0 & \\ 0 & 1 & 0 & & & & \\ 0 & 0 & 1 & & & & \\ \hline & & & 0 & & & \\ & & & & 0 & & \\ & & & & & 0 & \\ & & & & & & \ddots \end{array}\right).$$

Thus we can associate with each distinct root $\lambda_i$ a *projection operator* $\mathbf{P}_i$,

$$\mathbf{P}_i = \prod_{j\neq i}\frac{\mathbf{M} - \lambda_j\mathbf{1}}{\lambda_i - \lambda_j}\,, \tag{H.17}$$

which acts as identity on the $i$th subspace, and zero elsewhere. For example, the projection operator onto the $\lambda_1$ subspace is

$$\mathbf{P}_1 = \mathbf{C}^\dagger\left(\begin{array}{ccc|cccc} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ \hline & & & 0 & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{array}\right)\mathbf{C}\,. \tag{H.18}$$

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 $\mathbf{P}_i$: The matrices $\mathbf{P}_i$ are *orthogonal*

$$\mathbf{P}_i\mathbf{P}_j = \delta_{ij}\mathbf{P}_j\,, \qquad \text{(no sum on } j\text{)}\,, \tag{H.19}$$

and satisfy the *completeness relation*

$$\sum_{i=1}^{r}\mathbf{P}_i = \mathbf{1}\,. \tag{H.20}$$

As $\mathrm{tr}\,(\mathbf{C}\mathbf{P}_i\mathbf{C}^\dagger) = \mathrm{tr}\,\mathbf{P}_i$, the dimension of the $i$th subspace is given by

$$d_i = \mathrm{tr}\,\mathbf{P}_i\,. \tag{H.21}$$

It follows from the characteristic equation (H.16) and the form of the projection operator (H.17) that $\lambda_i$ is the eigenvalue of $\mathbf{M}$ on $\mathbf{P}_i$ subspace:

$$\mathbf{M}\mathbf{P}_i = \lambda_i \mathbf{P}_i \,, \qquad \text{(no sum on } i\text{)}\,. \tag{H.22}$$

Hence, any matrix polynomial $f(\mathbf{M})$ takes the scalar value $f(\lambda_i)$ on the $\mathbf{P}_i$ subspace

$$f(\mathbf{M})\mathbf{P}_i = f(\lambda_i)\mathbf{P}_i \,. \tag{H.23}$$

This, of course, is the reason why one wants to work with irreducible reps: they reduce matrices and "operators" to pure numbers.

### H.2.2 Irreducible representations

Suppose there exist several linearly independent invariant $[d\times d]$ hermitian matrices $\mathbf{M}_1, \mathbf{M}_2, \ldots$, and that we have used $\mathbf{M}_1$ to decompose the $d$-dimensional vector space $V = V_1 \oplus V_2 \oplus \cdots$. Can $\mathbf{M}_2, \mathbf{M}_3, \ldots$ be used to further decompose $V_i$? Further decomposition is possible if, and only if, the invariant matrices commute:

$$[\mathbf{M}_1, \mathbf{M}_2] = 0 \,, \tag{H.24}$$

or, equivalently, if projection operators $\mathbf{P}_j$ constructed from $\mathbf{M}_2$ commute with projection operators $\mathbf{P}_i$ constructed from $\mathbf{M}_1$,

$$\mathbf{P}_i \mathbf{P}_j = \mathbf{P}_j \mathbf{P}_i \,. \tag{H.25}$$

Usually the simplest choices of independent invariant matrices do not commute. In that case, the projection operators $\mathbf{P}_i$ constructed from $\mathbf{M}_1$ can be used to project commuting pieces of $\mathbf{M}_2$:

$$\mathbf{M}_2^{(i)} = \mathbf{P}_i \mathbf{M}_2 \mathbf{P}_i \,, \qquad \text{(no sum on } i\text{)}\,.$$

That $\mathbf{M}_2^{(i)}$ commutes with $\mathbf{M}_1$ follows from the orthogonality of $\mathbf{P}_i$:

$$[\mathbf{M}_2^{(i)}, \mathbf{M}_1] = \sum_j \lambda_j [\mathbf{M}_2^{(i)}, \mathbf{P}_j] = 0 \,. \tag{H.26}$$

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 (H.15) 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 (H.13), an invariant matrix $\mathbf{M}$ commutes with group transformations $[G, \mathbf{M}] = 0$. Projection operators (H.17) constructed from $\mathbf{M}$ are polynomials in $\mathbf{M}$, so they also commute with all $g \in \mathcal{G}$:

$$[G, \mathbf{P}_i] = 0 \tag{H.27}$$

Hence, a $[d \times d]$ matrix rep can be written as a direct sum of $[d_i \times d_i]$ matrix reps:

$$G = \mathbf{1}G\mathbf{1} = \sum_{i,j} \mathbf{P}_i G \mathbf{P}_j = \sum_i \mathbf{P}_i G \mathbf{P}_i = \sum_i G_i \,. \tag{H.28}$$

In the diagonalized rep (H.18), the matrix $\mathbf{g}$ has a block diagonal form:

$$\mathbf{C}\mathbf{g}\mathbf{C}^\dagger = \begin{bmatrix} \mathbf{g}_1 & 0 & 0 \\ 0 & \mathbf{g}_2 & 0 \\ 0 & 0 & \ddots \end{bmatrix}, \qquad \mathbf{g} = \sum_i \mathbf{C}^i \mathbf{g}_i \mathbf{C}_i \,. \tag{H.29}$$

The rep $\mathbf{g}_i$ acts only on the $d_i$-dimensional subspace $V_i$ consisting of vectors $\mathbf{P}_i q$, $q \in V$. 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 $V_i$:

$$V \overset{\mathbf{M}}{\to} V_1 \oplus V_2 \oplus \ldots \oplus V_r \,. \tag{H.30}$$

## H.3 Lattice derivatives

Consider a smooth function $\phi(x)$ evaluated on a finite $d$-dimensional lattice

$$\phi_\ell = \phi(x)\,, \qquad x = a\ell = \text{lattice point}\,, \quad \ell \in \mathbf{Z}^d\,, \tag{H.31}$$

where $a$ is the lattice spacing and there are $N^d$ points in all. A vector $\phi$ specifies a lattice configuration. Assume the lattice is hyper-cubic, and let $\hat{n}_\mu \in \{\hat{n}_1, \hat{n}_2, \cdots, \hat{n}_d\}$ be the unit lattice cell vectors pointing along the $d$ positive directions, $|\hat{n}_\mu| = 1$ . The *lattice partial derivative* is then

$$(\partial_\mu \phi)_\ell = \frac{\phi(x + a\hat{n}_\mu) - \phi(x)}{a} = \frac{\phi_{\ell+\hat{n}_\mu} - \phi_\ell}{a} \,.$$

Anything else with the correct $a \to 0$ limit would do, but this is the simplest choice. We can rewrite the derivative as a linear operator, by introducing the *hopping operator* (or "shift," or "step") in the direction $\mu$

$$\left(\mathbf{h}_\mu\right)_{\ell j} = \delta_{\ell+\hat{n}_\mu, j} \,. \tag{H.32}$$

As $\mathbf{h}$ will play a central role in what follows, it pays to understand what it does, so we write it out for the 1-dimensional case in its full $[N\times N]$ matrix glory:

$$\mathbf{h} = \begin{pmatrix} 0 & 1 & & & & \\ & 0 & 1 & & & \\ & & 0 & 1 & & \\ & & & & \ddots & \\ & & & & 0 & 1 \\ 1 & & & & & 0 \end{pmatrix} . \tag{H.33}$$

We will assume throughout that the lattice is *periodic* in each $\hat{n}_\mu$ direction; this is the easiest boundary condition to work with if we are interested in large lattices where surface effects are negligible.

Applied on the lattice configuration $\phi = (\phi_1, \phi_2, \cdots, \phi_N)$, the hopping operator shifts the lattice by one site, $\mathbf{h}\phi = (\phi_2, \phi_3, \cdots, \phi_N, \phi_1)$. Its transpose shifts the entries the other way, so the transpose is also the inverse

$$\mathbf{h}^{-1} = \mathbf{h}^T \,. \tag{H.34}$$

The lattice derivative can now be written as a multiplication by a matrix:

$$\partial_\mu \phi_\ell = \frac{1}{a}\left(\mathbf{h}_\mu - \mathbf{1}\right)_{\ell j} \phi_j \,.$$

In the 1-dimensional case the $[N\times N]$ matrix representation of the lattice derivative is:

$$\partial = \frac{1}{a}\begin{pmatrix} -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & -1 & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \\ 1 & & & & & -1 \end{pmatrix} . \tag{H.35}$$

To belabor the obvious: On a finite lattice of $N$ points a derivative is simply a finite $[N\times N]$ 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.

### H.3.1 Lattice Laplacian

In order to get rid of some of the lattice indices it is convenient to employ vector notation for the terms bilinear in $\phi$, and keep the rest lumped into "interaction,"

$$S[\phi] = -\frac{M^2}{2}\phi^T\cdot\phi - \frac{C}{2}\left[\left(\mathbf{h}_\mu - \mathbf{1}\right)\phi\right]^T\cdot\left(\mathbf{h}_\mu - \mathbf{1}\right)\phi + S_I[\phi]\,. \tag{H.36}$$

For example, for the discretized Landau Hamiltonian $M^2/2 = \beta m_0^2/2$, $C = \beta/a^2$, and the quartic term $S_I[\phi]$ is local site-by-site, $\gamma_{\ell_1\ell_2\ell_3\ell_4} = -4!\,\beta u\,\delta_{\ell_1\ell_2}\delta_{\ell_2\ell_3}\delta_{\ell_3\ell_4}$, so this general quartic coupling is a little bit of an overkill, but by the time we get to the Fourier-transformed theory, it will make sense as a momentum conserving vertex (H.62).

In the continuum integration by parts moves $\partial_\mu$ around; on a lattice this amounts to a matrix transposition

$$\left[\left(\mathbf{h}_\mu - \mathbf{1}\right)\phi\right]^T\cdot\left[\left(\mathbf{h}_\mu - \mathbf{1}\right)\phi\right] = \phi^T\cdot(\mathbf{h}_\mu^{-1} - \mathbf{1})\left(\mathbf{h}_\mu - \mathbf{1}\right)\cdot\phi\,.$$

If you are wondering where the "integration by parts" minus sign is, it is there in discrete case at well. It comes from the identity $\partial^T = -\mathbf{h}^{-1}\partial$. The combination $\Delta = \mathbf{h}^{-1}\partial^2$

$$\Delta = -\frac{1}{a^2}\sum_{\mu=1}^{d}(\mathbf{h}_\mu^{-1} - \mathbf{1})\left(\mathbf{h}_\mu - \mathbf{1}\right) = -\frac{2}{a^2}\sum_{\mu=1}^{d}\left(\mathbf{1} - \frac{1}{2}(\mathbf{h}_\mu^{-1} + \mathbf{h}_\mu)\right) \tag{H.37}$$

is the *lattice Laplacian*. We shall show below that this Laplacian has the correct continuum limit. It is the simplest spatial derivative allowed for $x \to -x$ symmetric actions. In the 1-dimensional case the $[N\times N]$ matrix representation of the lattice Laplacian is:

$$\Delta = \frac{1}{a^2}\begin{pmatrix} -2 & 1 & & & & 1 \\ 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & 1 & & \ddots & \\ & & & & & 1 \\ 1 & & & & 1 & -2 \end{pmatrix}. \tag{H.38}$$

The lattice Laplacian measures the second variation of a field $\phi_\ell$ across three neighboring sites. 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, $\phi_\ell = \phi(\ell)$, where $\phi(\ell)$ is defined by the value of the continuum function $\phi(x) = x^2$ at the lattice point $\ell$.

### H.3.2 Inverting the Laplacian

Evaluation of perturbative corrections in (26.21) requires that we come to grips with the "free" or "bare" propagator $M$ . While the the Laplacian is a simple difference operator (H.38), its inverse is a messier object. A way to compute is to start expanding $M$ as a power series in the Laplacian

$$\beta M = \frac{1}{m_0'^2\mathbf{1} - \Delta} = \frac{1}{m_0'^2}\sum_{k=0}^{\infty}\left(\frac{1}{m_0'^2}\right)^k \Delta^k . \tag{H.39}$$

As $\Delta$ is a finite matrix, the expansion is convergent for sufficiently large $m_0'^2$. To get a feeling for what is involved in evaluating such series, evaluate $\Delta^2$ in the 1-dimensional case:

$$\Delta^2 = \frac{1}{a^4}\begin{pmatrix} 6 & -4 & 1 & & & 1 & -4 \\ -4 & 6 & -4 & 1 & & & \\ 1 & -4 & 6 & -4 & 1 & & \\ & 1 & -4 & \ddots & & & \\ & & & & & 6 & -4 \\ -4 & 1 & & & 1 & -4 & 6 \end{pmatrix} . \tag{H.40}$$

What $\Delta^3$, $\Delta^4$, $\cdots$ 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. We will show how this works in sect. H.4.

## H.4 Periodic lattices

Our task now is to transform $M$ into a form suitable to evaluation of Feynman diagrams. The theory we will develop in this section is applicable only to *translationally invariant* saddle point configurations. bifurcation

Consider the effect of a $\phi \to \mathbf{h}\phi$ translation on the action

$$S[\mathbf{h}\phi] = -\frac{1}{2}\phi^T \cdot \mathbf{h}^T M^{-1}\mathbf{h} \cdot \phi - \frac{\beta g_0}{4!}\sum_{\ell=1}^{N^d}(\mathbf{h}\phi)_\ell^4 .$$

As $M^{-1}$ is constructed from $\mathbf{h}$ and its inverse, $M^{-1}$ and $\mathbf{h}$ commute, and the bilinear term is $\mathbf{h}$ invariant. In the quartic term $\mathbf{h}$ permutes cyclically the terms in the

sum, so the total action is translationally invariant

$$S[\mathbf{h}\phi] = S[\phi] = -\frac{1}{2}\phi^T \cdot M^{-1} \cdot \phi - \frac{\beta g_0}{4!}\sum_{\ell=1}^{N^d}\phi_\ell^4 \,. \tag{H.41}$$

If a function (in this case, the action $S[\phi]$) defined on a vector space (in this case, the configuration $\phi$) commutes with a linear operator $\mathbf{h}$, then the eigenvalues of $\mathbf{h}$ can be used to decompose the $\phi$ vector space into invariant subspaces. For a hyper-cubic lattice the translations in different directions commute, $\mathbf{h}_\mu\mathbf{h}_\nu = \mathbf{h}_\nu\mathbf{h}_\mu$, so it is sufficient to understand the spectrum of the 1-dimensional shift operator (H.33). To develop a feeling for how this reduction to invariant subspaces works in practice, let us continue humbly, by expanding the scope of our deliberations to a lattice consisting of 2 points.

### H.4.1 A 2-point lattice diagonalized

The action of the shift operator $\mathbf{h}$ (H.33) on a 2-point lattice $\phi = (\phi_1, \phi_2)$ is to permute the two lattice sites

$$\mathbf{h} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} .$$

As exchange repeated twice brings us back to the original configuration, $\mathbf{h}^2 = \mathbf{1}$, and the characteristic polynomial of $\mathbf{h}$ is

$$(\mathbf{h}+1)(\mathbf{h}-1) = 0 \,,$$

with eigenvalues $\lambda_0 = 1, \lambda_1 = -1$. Construct now the symmetrization, antisymmetrization projection operators

$$P_0 = \frac{\mathbf{h} - \lambda_1\mathbf{1}}{\lambda_0 - \lambda_1} = \frac{1}{2}(\mathbf{1} + \mathbf{h}) = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \tag{H.42}$$

$$P_1 = \frac{\mathbf{h} - \mathbf{1}}{-1-1} = \frac{1}{2}(\mathbf{1} - \mathbf{h}) = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} . \tag{H.43}$$

Noting that $P_0 + P_1 = \mathbf{1}$, we can project the lattice configuration $\phi$ onto the two eigenvectors of $\mathbf{h}$:

$$\phi = \mathbf{1}\phi = P_0 \cdot \phi + P_1 \cdot \phi \,,$$

$$\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \frac{(\phi_1+\phi_2)}{\sqrt{2}}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{(\phi_1-\phi_2)}{\sqrt{2}}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{H.44}$$

$$= \tilde{\phi}_0\hat{n}_0 + \tilde{\phi}_1\hat{n}_1 \,. \tag{H.45}$$

As $P_0P_1 = 0$, the symmetric and the antisymmetric configurations transform separately under any linear transformation constructed from $\mathbf{h}$ and its powers.

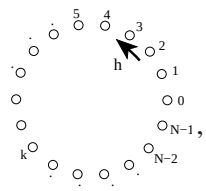
In this way the characteristic equation $\mathbf{h}^2 = \mathbf{1}$ enables us to reduce the 2-dimenional lattice configuration to two 1-dimensional ones, on which the value of the shift operator (shift matrix) $\mathbf{h}$ is a number, $\lambda \in \{1, -1\}$, and the eigenvectors are $\hat{n}_0 = \frac{1}{\sqrt{2}}(1, 1)$, $\hat{n}_1 = \frac{1}{\sqrt{2}}(1, -1)$. We have inserted $\sqrt{2}$ factors only for convenience, in order that the eigenvectors be normalized unit vectors. As we shall now see, $(\tilde{\phi}_0, \tilde{\phi}_1)$ is the 2-site periodic lattice discrete Fourier transform of the field $(\phi_1, \phi_2)$.

## H.5 Discrete Fourier transforms

Now let us generalize this reduction to a 1-dimensional periodic lattice with $N$ sites.

Each application of $\mathbf{h}$ translates the lattice one step; in $N$ steps the lattice is back in the original configuration

$$\mathbf{h}^N = \mathbf{1}$$




so the eigenvalues of $\mathbf{h}$ are the $N$ distinct $N$-th roots of unity

$$\mathbf{h}^N - \mathbf{1} = \prod_{k=0}^{N-1} (\mathbf{h} - \omega^k \mathbf{1}) = 0\,, \qquad \omega = e^{i\frac{2\pi}{N}}\,. \tag{H.46}$$

As the eigenvalues are all distinct and $N$ in number, the space is decomposed into $N$ 1-dimensional subspaces. The general theory (expounded in appendix H.2) associates with the $k$-th eigenvalue of $\mathbf{h}$ a projection operator that projects a configuration $\phi$ onto $k$-th eigenvector of $\mathbf{h}$,

$$P_k = \prod_{j \neq k} \frac{\mathbf{h} - \lambda_j \mathbf{1}}{\lambda_k - \lambda_j}\,. \tag{H.47}$$

A factor $(\mathbf{h} - \lambda_j \mathbf{1})$ kills the $j$-th eigenvector $\varphi_j$ component of an arbitrary vector in expansion $\phi = \cdots + \tilde{\phi}_j \varphi_j + \cdots$. The above product kills everything but the eigendirection $\varphi_k$, and the factor $\prod_{j \neq k}(\lambda_k - \lambda_j)$ ensures that $P_k$ is normalized as a projection operator. The set of the projection operators is complete

$$\sum_k P_k = \mathbf{1} \tag{H.48}$$

and orthonormal

$$P_k P_j = \delta_{kj} P_k \qquad \text{(no sum on } k\text{)}\,. \tag{H.49}$$

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 projection operators [1]. However, in case at hand the eigenvectors are so simple that we can forget the general theory, and construct the solutions of the eigenvalue condition

$$\mathbf{h}\,\varphi_k \quad = \quad \omega^k \varphi_k \tag{H.50}$$

by hand:

$$\frac{1}{\sqrt{N}}\begin{pmatrix} 0 & 1 & & & & \\ & 0 & 1 & & & \\ & & 0 & 1 & & \\ & & & & \ddots & \\ & & & & 0 & 1 \\ 1 & & & & & 0 \end{pmatrix}\begin{pmatrix} 1 \\ \omega^k \\ \omega^{2k} \\ \omega^{3k} \\ \vdots \\ \omega^{(N-1)k} \end{pmatrix} = \omega^k \frac{1}{\sqrt{N}}\begin{pmatrix} 1 \\ \omega^k \\ \omega^{2k} \\ \omega^{3k} \\ \vdots \\ \omega^{(N-1)k} \end{pmatrix}$$

The $1/\sqrt{N}$ factor is chosen in order that $\varphi_k$ be normalized unit vectors

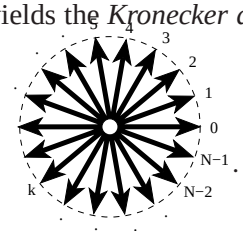
$$\begin{aligned} \varphi_k^\dagger \cdot \varphi_k &= \frac{1}{N}\sum_{k=0}^{N-1} 1 \;=\; 1\,, \qquad \text{(no sum on } k\text{)} \\ \varphi_k^\dagger &= \frac{1}{\sqrt{N}}\left(1, \omega^{-k}, \omega^{-2k}, \cdots, \omega^{-(N-1)k}\right)\,. \end{aligned} \tag{H.51}$$

The eigenvectors are orthonormal

$$\varphi_k^\dagger \cdot \varphi_j = \delta_{kj}\,, \tag{H.52}$$

as the explicit evaluation of $\varphi_k^\dagger \cdot \varphi_j$ yields the *Kronecker delta function for a periodic lattice*

$$\delta_{kj} = \frac{1}{N}\sum_{\ell=0}^{N-1} e^{i\frac{2\pi}{N}(k-j)\ell} \tag{H.53}$$




The sum is over the $N$ unit vectors pointing at a uniform distribution of points on the complex unit circle; they cancel each other unless $k = j \pmod N$, in which case each term in the sum equals 1.

The projection operators can be expressed in terms of the eigenvectors (H.50), (H.51) as

$$(P_k)_{\ell\ell'} = (\varphi_k)_\ell (\varphi_k^\dagger)_{\ell'} = \frac{1}{N} e^{i\frac{2\pi}{N}(\ell-\ell')k}\,, \qquad \text{(no sum on } k\text{)}\,. \tag{H.54}$$

The completeness (H.48) follows from (H.53), and the orthonormality (H.49) from (H.52).

$\tilde{\phi}_k$, the projection of the $\phi$ configuration on the $k$-th subspace is given by

$$\begin{aligned}(P_k \cdot \phi)_\ell &= \tilde{\phi}_k\,(\varphi_k)_\ell\,, \qquad \text{(no sum on } k\text{)}\\ \tilde{\phi}_k &= \varphi_k^\dagger \cdot \phi = \frac{1}{\sqrt{N}}\sum_{\ell=0}^{N-1} e^{-i\frac{2\pi}{N}k\ell}\phi_\ell \end{aligned} \tag{H.55}$$

We recognize $\tilde{\phi}_k$ as the *discrete Fourier transform* of $\phi_\ell$. 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 the generator of translations) and when are they the natural set of basis functions (only if the theory is translationally invariant).

### H.5.1 Fourier transform of the propagator

Now insert the identity $\sum P_k = \mathbf{1}$ wherever profitable:

$$\mathbf{M} = \mathbf{1M1} = \sum_{kk'} P_k \mathbf{M} P_{k'} = \sum_{kk'} \varphi_k(\varphi_k^\dagger \cdot \mathbf{M} \cdot \varphi_{k'})\varphi_{k'}^\dagger\,.$$

The matrix

$$\tilde{M}_{kk'} = (\varphi_k^\dagger \cdot \mathbf{M} \cdot \varphi_{k'}) \tag{H.56}$$

is the Fourier space representation of $\mathbf{M}$. No need to stop here - the terms in the action (H.41) that couple four (and, in general, 3, 4, $\cdots$) fields also have the Fourier space representations

$$\begin{aligned}\gamma_{\ell_1\ell_2\cdots\ell_n}\,\phi_{\ell_1}\phi_{\ell_2}\cdots\phi_{\ell_n} &= \tilde{\gamma}_{k_1k_2\cdots k_n}\,\tilde{\phi}_{k_1}\tilde{\phi}_{k_2}\cdots\tilde{\phi}_{k_n}\,,\\ \tilde{\gamma}_{k_1k_2\cdots k_n} &= \gamma_{\ell_1\ell_2\cdots\ell_n}(\varphi_{k_1})_{\ell_1}(\varphi_{k_2})_{\ell_2}\cdots(\varphi_{k_n})_{\ell_n}\\ &= \frac{1}{N^{n/2}}\sum_{\ell_1\cdots\ell_n}\gamma_{\ell_1\ell_2\cdots\ell_n}\,e^{-i\frac{2\pi}{N}(k_1\ell_1+\cdots+k_n\ell_n)}\,.\end{aligned} \tag{H.57}$$

According to (H.52) the matrix $U_{k\ell} = (\varphi_k)_\ell = \frac{1}{\sqrt{N}}e^{i\frac{2\pi}{N}k\ell}$ is a unitary matrix, and the Fourier transform is a linear, unitary transformation $UU^\dagger = \sum P_k = \mathbf{1}$ with Jacobian det $U = 1$. The form of the action (H.41) does not change under $\phi \to \tilde{\phi}_k$ transformation, and from the formal point of view, 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 $\mathbf{M}$ is the trace in either representation

$$\begin{aligned}\mathrm{tr}\,\mathbf{M} &= \sum_\ell M_{\ell\ell} = \sum_{kk'}\sum_\ell (P_k\mathbf{M}P_{k'})_{\ell\ell}\\ &= \sum_{kk'}\sum_\ell (\varphi_k)_\ell(\varphi_k^\dagger\cdot\mathbf{M}\cdot\varphi_{k'})(\varphi_{k'}^\dagger)_\ell = \sum_{kk'}\delta_{kk'}\tilde{M}_{kk'} = \mathrm{tr}\,\tilde{\mathbf{M}}\,.\end{aligned} \tag{H.58}$$

From this it follows that $\mathrm{tr}\,\mathbf{M}^n = \mathrm{tr}\,\tilde{\mathbf{M}}^n$, and from the tr ln = ln tr relation that $\det \mathbf{M} = \det \tilde{\mathbf{M}}$. In fact, any scalar combination of $\phi$'s, $J$'s and couplings, such as the partition function $Z[J]$, has exactly the same form in the configuration and the Fourier space.

OK, a dizzying quantity of indices. But what's the pay-back?

### H.5.2 Lattice Laplacian diagonalized

Now use the eigenvalue equation (H.50) to convert **h** matrices into scalars. If **M** *commutes* with **h**, then $(\varphi_k^\dagger \cdot \mathbf{M} \cdot \varphi_{k'}) = \tilde{M}_k \delta_{kk'}$, and the matrix **M** acts as a multiplication by the scalar $\tilde{M}_k$ on the $k$-th subspace. For example, for the 1-dimensional version of the lattice Laplacian (H.37) the projection on the $k$-th subspace is

$$\begin{aligned}(\varphi_k^\dagger \cdot \Delta \cdot \varphi_{k'}) &= \frac{2}{a^2}\left(\frac{1}{2}(\omega^{-k}+\omega^k)-1\right)(\varphi_k^\dagger \cdot \varphi_{k'}) \\ &= \frac{2}{a^2}\left(\cos\left(\frac{2\pi}{N}k\right)-1\right)\delta_{kk'} \end{aligned} \tag{H.59}$$

In the $k$-th subspace the bare propagator (H.59) is simply a number, and, in contrast to the mess generated by (H.39), there is nothing to inverting $M^{-1}$:

$$(\varphi_{\mathbf{k}}^\dagger \cdot M \cdot \varphi_{\mathbf{k}'}) = (\tilde{G}_0)_{\mathbf{k}}\delta_{\mathbf{k}\mathbf{k}'} = \frac{1}{\beta}\frac{\delta_{\mathbf{k}\mathbf{k}'}}{m_0'^2 - \frac{2c}{a^2}\sum_{\mu=1}^d\left(\cos\left(\frac{2\pi}{N}k_\mu\right)-1\right)}\,, \tag{H.60}$$

where $\mathbf{k} = (k_1, k_2, \cdots, k_\mu)$ is a $d$-dimensional vector in the $N^d$-dimensional dual lattice.

Going back to the partition function (26.21) and sticking in the factors of **1** into the bilinear part of the interaction, we replace the spatial $J_\ell$ by its Fourier transform $\tilde{J}_k$, and the spatial propagator $(M)_{\ell\ell'}$ by the diagonalized Fourier transformed $(\tilde{G}_0)_k$

$$J^T \cdot M \cdot J = \sum_{k,k'}(J^T \cdot \varphi_k)(\varphi_k^\dagger \cdot M \cdot \varphi_{k'})(\varphi_{k'}^\dagger \cdot J) = \sum_k \tilde{J}_k^\dagger (\tilde{G}_0)_k \tilde{J}_k\,. \tag{H.61}$$

What's the price? The interaction term $S_I[\phi]$ (which in (26.21) was local in the configuration space) now has a more challenging $k$ dependence in the Fourier transform version (H.57). For example, the locality of the quartic term leads to the 4-vertex *momentum conservation* in the Fourier space

$$\begin{aligned} S_I[\phi] &= \frac{1}{4!}\gamma_{\ell_1\ell_2\ell_3\ell_4}\,\phi_{\ell_1}\phi_{\ell_2}\phi_{\ell_3}\phi_{\ell_4} = -\beta u\sum_{\ell=1}^{N^d}(\phi_\ell)^4 \;\Rightarrow \\ &= -\beta u\frac{1}{N^{3d/2}}\sum_{\{\mathbf{k}_i\}}^{N}\delta_{0,\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3+\mathbf{k}_4}\,\tilde{\phi}_{\mathbf{k}_1}\tilde{\phi}_{\mathbf{k}_2}\tilde{\phi}_{\mathbf{k}_3}\tilde{\phi}_{\mathbf{k}_4}\,. \end{aligned} \tag{H.62}$$

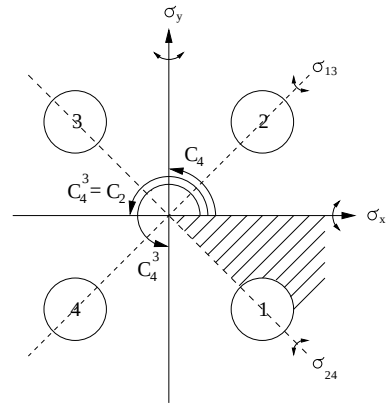



**Figure H.1:** Symmetries of four disks on a square. A fundamental domain indicated by the shaded wedge.

## H.6 $C_{4v}$ factorization

If an $N$-disk arrangement has $C_N$ symmetry, and the disk visitation sequence is given by disk labels $\{\epsilon_1\epsilon_2\epsilon_3\ldots\}$, only the relative increments $\rho_i = \epsilon_{i+1} - \epsilon_i \bmod N$ matter. Symmetries under reflections across axes increase the group to $C_{Nv}$ 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 H.1). The symmetry group consists of the identity **e**, the two reflections $\sigma_x$, $\sigma_y$ across $x$, $y$ axes, the two diagonal reflections $\sigma_{13}$, $\sigma_{24}$, and the three rotations $C_4$, $C_2$ and $C_4^3$ 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 → disk 2 segment as the starting segment, this symbol string is mapped into the disk visitation sequence $1_{+1}2_{+1}3_{+2}1\ldots = \overline{123}$, 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 $\overline{\underline{1}12}$ 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 | ↔ | $\underline{1}$ |
| 1234 | ↔ | 1 |
| 13 | ↔ | 2 |

Conversions for all periodic orbits of reduced symbol period less than 5 are listed in table **??**.

**Table H.1:** $C_{4v}$ 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 $C_{4v}$ 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. H.6 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 $\overline{2}$ is the sole boundary orbit, invariant both under a rotation by $\pi$ 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 $C_{4v}$ transformations.

| $\tilde{p}$ | $p$ | $\mathbf{h}_{\tilde{p}}$ |
|---|---|---|
| 0 | 1 2 | $\sigma_x$ |
| 1 | 1 2 3 4 | $C_4$ |
| 2 | 1 3 | $C_2$, $\sigma_{13}$ |
| 01 | 12 14 | $\sigma_{24}$ |
| 02 | 12 43 | $\sigma_y$ |
| 12 | 12 41 34 23 | $C_4^3$ |
| 001 | 121 232 343 414 | $C_4$ |
| 002 | 121 343 | $C_2$ |
| 011 | 121 434 | $\sigma_y$ |
| 012 | 121 323 | $\sigma_{13}$ |
| 021 | 124 324 | $\sigma_{13}$ |
| 022 | 124 213 | $\sigma_x$ |
| 112 | 123 | $e$ |
| 122 | 124 231 342 413 | $C_4$ |

| $\tilde{p}$ | $p$ | $\mathbf{h}_{\tilde{p}}$ |
|---|---|---|
| 0001 | 1212 1414 | $\sigma_{24}$ |
| 0002 | 1212 4343 | $\sigma_y$ |
| 0011 | 1212 3434 | $C_2$ |
| 0012 | 1212 4141 3434 2323 | $C_4^3$ |
| 0021 ($a$) | 1213 4142 3431 2324 | $C_4^3$ |
| 0022 | 1213 | $e$ |
| 0102 ($a$) | 1214 2321 3432 4143 | $C_4$ |
| 0111 | 1214 3234 | $\sigma_{13}$ |
| 0112 ($b$) | 1214 2123 | $\sigma_x$ |
| 0121 ($b$) | 1213 2124 | $\sigma_x$ |
| 0122 | 1213 1413 | $\sigma_{24}$ |
| 0211 | 1243 2134 | $\sigma_x$ |
| 0212 | 1243 1423 | $\sigma_{24}$ |
| 0221 | 1242 1424 | $\sigma_{24}$ |
| 0222 | 1242 4313 | $\sigma_y$ |
| 1112 | 1234 2341 3412 4123 | $C_4$ |
| 1122 | 1231 3413 | $C_2$ |
| 1222 | 1242 4131 3424 2313 | $C_4^3$ |

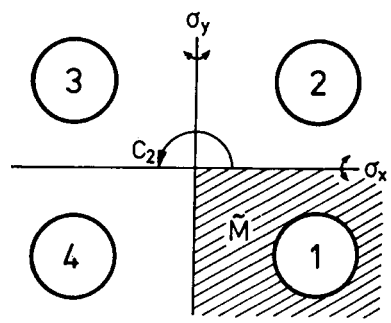



**Figure H.2:** Symmetries of four disks on a rectangle. A fundamental domain indicated by the shaded wedge.

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 $C_{4v}$ group elements $g_1 = C$, $g_2 = C^2$, $g_{\underline{1}} = \sigma_{diag}C = \sigma_{axis}$, where $C$ is a rotation by $\pi/2$. In the $\overline{\underline{1}12}$ example worked out above, this yields $g_{\underline{1}12} = g_2 g_1 g_{\underline{1}} = C^2 C\sigma_{axis} = \sigma_{diag}$, listed in the last column of table **??**. 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 $C_{4v}$ 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 degenerate pair of 2-dimensional representations $E$. Substituting the $C_{4v}$ characters

| $C_{4v}$ | $A_1$ | $A_2$ | $B_1$ | $B_2$ | $E$ |
|---|---|---|---|---|---|
| $e$ | 1 | 1 | 1 | 1 | 2 |
| $C_2$ | 1 | 1 | 1 | 1 | -2 |
| $C_4, C_4^3$ | 1 | 1 | -1 | -1 | 0 |
| $\sigma_{axes}$ | 1 | -1 | 1 | -1 | 0 |
| $\sigma_{diag}$ | 1 | -1 | -1 | 1 | 0 |

into (19.15) we obtain:

$$
\begin{array}{rlcccccc}
h_{\tilde{p}} & & & A_1 & A_2 & B_1 & B_2 & E \\
e: & (1-t_{\tilde{p}})^8 & = & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}})^4 \\
C_2: & (1-t_{\tilde{p}}^2)^4 & = & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}})^4 \\
C_4, C_4^3: & (1-t_{\tilde{p}}^4)^2 & = & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1+t_{\tilde{p}}^2)^2 \\
\sigma_{axes}: & (1-t_{\tilde{p}}^2)^4 & = & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1-t_{\tilde{p}}^2)^2 \\
\sigma_{diag}: & (1-t_{\tilde{p}}^2)^4 & = & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}^2)^2
\end{array}
$$

The possible irreducible segment group elements $\mathbf{h}_{\tilde{p}}$ are listed in the first column; $\sigma_{axes}$ denotes a reflection across either the x-axis or the y-axis, and $\sigma_{diag}$ denotes a reflection across a diagonal (see figure H.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 $\mathbf{h}_p = (C_2 + \sigma_x)/2$ (axes) and $\mathbf{h}_p = (C_2 + \sigma_{13})/2$ (diagonals) respectively:

$$
\begin{array}{rlcl}
 & & & A_1 \quad A_2 \quad B_1 \quad B_2 \quad E \\
\text{axes:} & (1-t_{\tilde{p}}^2)^2 & = & (1-t_{\tilde{p}})(1-0t_{\tilde{p}})(1-t_{\tilde{p}})(1-0t_{\tilde{p}})(1+t_{\tilde{p}})^2 \\
\text{diagonals:} & (1-t_{\tilde{p}}^2)^2 & = & (1-t_{\tilde{p}})(1-0t_{\tilde{p}})(1-0t_{\tilde{p}})(1-t_{\tilde{p}})(1+t_{\tilde{p}})^2
\end{array}
\tag{H.63}
$$

(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 $\overline{13}$, $\overline{24}$ occur; they correspond to the $\overline{2}$ fixed point in the fundamental domain.

The $A_1$ subspace in $C_{4v}$ cycle expansion is given by

$$
\begin{aligned}
1/\zeta_{A_1} &= (1-t_0)(1-t_1)(1-t_2)(1-t_{01})(1-t_{02})(1-t_{12}) \\
&\quad (1-t_{001})(1-t_{002})(1-t_{011})(1-t_{012})(1-t_{021})(1-t_{022})(1-t_{112}) \\
&\quad (1-t_{122})(1-t_{0001})(1-t_{0002})(1-t_{0011})(1-t_{0012})(1-t_{0021})\ldots \\
&= 1-t_0-t_1-t_2-(t_{01}-t_0t_1)-(t_{02}-t_0t_2)-(t_{12}-t_1t_2) \\
&\quad -(t_{001}-t_0t_{01})-(t_{002}-t_0t_{02})-(t_{011}-t_1t_{01}) \\
&\quad -(t_{022}-t_2t_{02})-(t_{112}-t_1t_{12})-(t_{122}-t_2t_{12}) \\
&\quad -(t_{012}+t_{021}+t_0t_1t_2-t_0t_{12}-t_1t_{02}-t_2t_{01})\ldots
\end{aligned}
\tag{H.64}
$$

(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 $\tilde{p}$ symbol string. For $B_2$ all $t_p$ with an odd total number of 0's and 1's change sign:

$$
\begin{aligned}
1/\zeta_{B_2} &= (1+t_0)(1+t_1)(1-t_2)(1-t_{01})(1+t_{02})(1+t_{12}) \\
&\quad (1+t_{001})(1-t_{002})(1+t_{011})(1-t_{012})(1-t_{021})(1+t_{022})(1-t_{112}) \\
&\quad (1+t_{122})(1-t_{0001})(1+t_{0002})(1-t_{0011})(1+t_{0012})(1+t_{0021})\ldots \\
&= 1+t_0+t_1-t_2-(t_{01}-t_0t_1)+(t_{02}-t_0t_2)+(t_{12}-t_1t_2) \\
&\quad +(t_{001}-t_0t_{01})-(t_{002}-t_0t_{02})+(t_{011}-t_1t_{01}) \\
&\quad +(t_{022}-t_2t_{02})-(t_{112}-t_1t_{12})+(t_{122}-t_2t_{12}) \\
&\quad -(t_{012}+t_{021}+t_0t_1t_2-t_0t_{12}-t_1t_{02}-t_2t_{01})\ldots
\end{aligned}
\tag{H.65}
$$

The form of the remaining cycle expansions depends crucially on the special role played by the boundary orbits: by (H.63) the orbit $t_2$ does not contribute to $A_2$ and $B_1$,

$$
\begin{aligned}
1/\zeta_{A_2} &= (1+t_0)(1-t_1)(1+t_{01})(1+t_{02})(1-t_{12}) \\
&\quad (1-t_{001})(1-t_{002})(1+t_{011})(1+t_{012})(1+t_{021})(1+t_{022})(1-t_{112}) \\
&\quad (1-t_{122})(1+t_{0001})(1+t_{0002})(1-t_{0011})(1-t_{0012})(1-t_{0021})\ldots \\
&= 1+t_0-t_1+(t_{01}-t_0t_1)+t_{02}-t_{12} \\
&\quad -(t_{001}-t_0t_{01})-(t_{002}-t_0t_{02})+(t_{011}-t_1t_{01}) \\
&\quad +t_{022}-t_{122}-(t_{112}-t_1t_{12})+(t_{012}+t_{021}-t_0t_{12}-t_1t_{02})\ldots
\end{aligned}
\tag{H.66}
$$

and

$$
\begin{aligned}
1/\zeta_{B_1} &= (1-t_0)(1+t_1)(1+t_{01})(1-t_{02})(1+t_{12}) \\
&\quad (1+t_{001})(1-t_{002})(1-t_{011})(1+t_{012})(1+t_{021})(1-t_{022})(1-t_{112}) \\
&\quad (1+t_{122})(1+t_{0001})(1-t_{0002})(1-t_{0011})(1+t_{0012})(1+t_{0021})\ldots \\
&= 1-t_0+t_1+(t_{01}-t_0t_1)-t_{02}+t_{12} \\
&\quad +(t_{001}-t_0t_{01})-(t_{002}-t_0t_{02})-(t_{011}-t_1t_{01}) \\
&\quad -t_{022}+t_{122}-(t_{112}-t_1t_{12})+(t_{012}+t_{021}-t_0t_{12}-t_1t_{02})\ldots
\end{aligned}
\tag{H.67}
$$

In the above we have assumed that $t_2$ does not change sign under $C_{4v}$ reflections. For the mixed-symmetry subspace $E$ the curvature expansion is given by

$$
\begin{aligned}
1/\zeta_E &= 1+t_2+(-{t_0}^2+{t_1}^2)+(2t_{002}-t_2{t_0}^2-2t_{112}+t_2{t_1}^2) \\
&\quad +(2t_{0011}-2t_{0022}+2t_2t_{002}-{t_{01}}^2-{t_{02}}^2+2t_{1122}-2t_2t_{112} \\
&\quad +{t_{12}}^2-{t_0}^2{t_1}^2)+(2t_{00002}-2t_{00112}+2t_2t_{0011}-2t_{00121}-2t_{00211} \\
&\quad +2t_{00222}-2t_2t_{0022}+2t_{01012}+2t_{01021}-2t_{01102}-t_2{t_{01}}^2+2t_{02022} \\
&\quad -t_2{t_{02}}^2+2t_{11112}-2t_{11222}+2t_2t_{1122}-2t_{12122}+t_2{t_{12}}^2-t_2{t_0}^2{t_1}^2 \\
&\quad +2t_{002}(-{t_0}^2+{t_1}^2)-2t_{112}(-{t_0}^2+{t_1}^2))
\end{aligned}
\tag{H.68}
$$

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 (13.40).

[exercise 18.9]

## H.7 $C_{2v}$ factorization

An arrangement of four identical disks on the vertices of a rectangle has $C_{2v}$ symmetry (figure H.2b). $C_{2v}$ consists of $\{e, \sigma_x, \sigma_y, C_2\}$, i.e., the reflections across the symmetry axes and a rotation by $\pi$.

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 $g_0 = \sigma_x$ and $g_1 = \sigma_y$ 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 $C_2 = \sigma_x\sigma_y$, to which we assign the label 2. For example, a ternary string 0 0 1 0 2 0 1 . . . is converted into 12143123. . ., and the associated group-theory weight is given by $\ldots g_1 g_0 g_2 g_0 g_1 g_0 g_0$.

Short ternary cycles and the corresponding 4-disk cycles are listed in table **??**. 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 $C_{2v}$ symmetries.

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 $C_{2v}$ 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

| $C_{2v}$ | $A_1$ | $A_2$ | $B_1$ | $B_2$ |
|---|---|---|---|---|
| $e$ | 1 | 1 | 1 | 1 |
| $C_2$ | 1 | 1 | −1 | −1 |
| $\sigma_x$ | 1 | −1 | 1 | −1 |
| $\sigma_y$ | 1 | −1 | −1 | 1 |

**Table H.2:** $C_{2v}$ 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 $C_{2v}$ 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 $C_{2v}$ transformations. The full space orbit listed here is generated from the symmetry reduced code by the rules given in sect. H.7, starting from disk 1.

| $\tilde{p}$ | $p$ | **g** | $\tilde{p}$ | $p$ | **g** |
|---|---|---|---|---|---|
| 0 | 1 4 | $\sigma_y$ | 0001 | 1414 3232 | $C_2$ |
| 1 | 1 2 | $\sigma_x$ | 0002 | 1414 2323 | $\sigma_x$ |
| 2 | 1 3 | $C_2$ | 0011 | 1412 | $e$ |
| 01 | 14 32 | $C_2$ | 0012 | 1412 4143 | $\sigma_y$ |
| 02 | 14 23 | $\sigma_x$ | 0021 | 1413 4142 | $\sigma_y$ |
| 12 | 12 43 | $\sigma_y$ | 0022 | 1413 | $e$ |
| 001 | 141 232 | $\sigma_x$ | 0102 | 1432 4123 | $\sigma_y$ |
| 002 | 141 323 | $C_2$ | 0111 | 1434 3212 | $C_2$ |
| 011 | 143 412 | $\sigma_y$ | 0112 | 1434 2343 | $\sigma_x$ |
| 012 | 143 | $e$ | 0121 | 1431 2342 | $\sigma_x$ |
| 021 | 142 | $e$ | 0122 | 1431 3213 | $C_2$ |
| 022 | 142 413 | $\sigma_y$ | 0211 | 1421 2312 | $\sigma_x$ |
| 112 | 121 343 | $C_2$ | 0212 | 1421 3243 | $C_2$ |
| 122 | 124 213 | $\sigma_x$ | 0221 | 1424 3242 | $C_2$ |
| | | | 0222 | 1424 2313 | $\sigma_x$ |
| | | | 1112 | 1212 4343 | $\sigma_y$ |
| | | | 1122 | 1213 | $e$ |
| | | | 1222 | 1242 4313 | $\sigma_y$ |

Substituted into the factorized determinant (19.14), the contributions of periodic orbits split as follows

$$
\begin{array}{rlcccccc}
g_{\tilde{p}} & & & A_1 & A_2 & B_1 & B_2 \\
e: & (1-t_{\tilde{p}})^4 & = & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) \\
C_2: & (1-t_{\tilde{p}}^2)^2 & = & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) \\
\sigma_x: & (1-t_{\tilde{p}}^2)^2 & = & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) \\
\sigma_y: & (1-t_{\tilde{p}}^2)^2 & = & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1-t_{\tilde{p}})
\end{array}
$$

Cycle expansions follow by substituting cycles and their group theory factors from table **??**. For $A_1$ all characters are +1, and the corresponding cycle expansion is given in (H.64). Similarly, the totally antisymmetric subspace factorization $A_2$ is given by (H.65), the $B_2$ factorization of $C_{4v}$. For $B_1$ all $t_p$ with an odd total number of 0's and 2's change sign:

$$
\begin{aligned}
1/\zeta_{B_1} &= (1+t_0)(1-t_1)(1+t_2)(1+t_{01})(1-t_{02})(1+t_{12}) \\
&\quad (1-t_{001})(1+t_{002})(1+t_{011})(1-t_{012})(1-t_{021})(1+t_{022})(1+t_{112}) \\
&\quad (1-t_{122})(1+t_{0001})(1-t_{0002})(1-t_{0011})(1+t_{0012})(1+t_{0021})\ldots \\
&= 1+t_0-t_1+t_2+(t_{01}-t_0t_1)-(t_{02}-t_0t_2)+(t_{12}-t_1t_2) \\
&\quad -(t_{001}-t_0t_{01})+(t_{002}-t_0t_{02})+(t_{011}-t_1t_{01}) \\
&\quad +(t_{022}-t_2t_{02})+(t_{112}-t_1t_{12})-(t_{122}-t_2t_{12}) \\
&\quad -(t_{012}+t_{021}+t_0t_1t_2-t_0t_{12}-t_1t_{02}-t_2t_{01})\ldots
\end{aligned}
\tag{H.69}
$$

For $B_2$ all $t_p$ with an odd total number of 1's and 2's change sign:

$$\begin{aligned}
1/\zeta_{B_2} &= (1-t_0)(1+t_1)(1+t_2)(1+t_{01})(1+t_{02})(1-t_{12}) \\
&\quad (1+t_{001})(1+t_{002})(1-t_{011})(1-t_{012})(1-t_{021})(1-t_{022})(1+t_{112}) \\
&\quad (1+t_{122})(1+t_{0001})(1+t_{0002})(1-t_{0011})(1-t_{0012})(1-t_{0021})\dots \\
&= 1-t_0+t_1+t_2+(t_{01}-t_0t_1)+(t_{02}-t_0t_2)-(t_{12}-t_1t_2) \\
&\quad +(t_{001}-t_0t_{01})+(t_{002}-t_0t_{02})-(t_{011}-t_1t_{01}) \\
&\quad -(t_{022}-t_2t_{02})+(t_{112}-t_1t_{12})+(t_{122}-t_2t_{12}) \\
&\quad -(t_{012}+t_{021}+t_0t_1t_2-t_0t_{12}-t_1t_{02}-t_2t_{01})\dots
\end{aligned} \tag{H.70}$$

Note that all of the above cycle expansions group long orbits together with their pseudoorbit shadows, so that the shadowing arguments for convergence still apply.

The topological polynomial factorizes as

$$\frac{1}{\zeta_{A_1}} = 1-3z \quad , \quad \frac{1}{\zeta_{A_2}} = \frac{1}{\zeta_{B_1}} = \frac{1}{\zeta_{B_2}} = 1+z,$$

consistent with the 4-disk factorization (13.40).

## H.8 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

$$x_{n-1} = -\frac{1}{b}(1-ax_n^2-x_{n+1})\,. \tag{H.71}$$

Hence the time reversal amounts to $b \to 1/b$, $a \to a/b^2$ symmetry in the parameter plane, together with $x \to -x/b$ 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 ,

$$x_{n-1} = 1-ax_n^2-x_{n+1}\,, \tag{H.72}$$

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 Poincaré return map for a Hamiltonian flow. For the orientation reversing $b=1$ case we have

$$x_{n-1} = 1-ax_n^2+x_{n+1}\,, \tag{H.73}$$

and the non–wandering set is symmetric across the $x_{n+1}=-x_n$ diagonal.

## Commentary

**Remark H.1** Literature This material is covered in any introduction to linear algebra [1, 2, 3] or group theory [11, 10]. The exposition given in sects. H.2.1 and H.2.2 is taken from refs. [6, 7, 1]. Who wrote this down first we do not know, but we like Harter's exposition [8, 9, 12] best. Harter's theory of class algebrasoffers a more elegant and systematic way of constructing the maximal set of commuting invariant matrices $\mathbf{M}_i$ than the sketch offered in this section.

**Remark H.2** Labeling conventions While there is a variety of labeling conventions [16, 8] for the reduced $C_{4v}$ 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 $C_{4v}$ group elements.

**Remark H.3** $C_{2v}$ symmetry $C_{2v}$ is the symmetry of several systems studied in the literature, such as the stadium billiard [10], and the 2-dimensional anisotropic Kepler potential [4].

## Exercises

H.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 H.1.)

From W.G. Harter [12]

H.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 $\mathbf{P}_+$ and $\mathbf{P}_-$, and show that the 3-pendulum system decomposes into a 1-$d$ subspace, with eigenvalue $(\omega^{(-)})^2 = a+b$, and a 2-$d$ 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 $\mathbf{P}_+, \mathbf{P}_-$ *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 [12]

H.3. **Laplacian is a non-local operator.**

While the Laplacian is a simple tri-diagonal difference operator (H.38), 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

$$\frac{1}{m^2\mathbf{1} - \Delta} = \frac{1}{m^2}\sum_{n=0}^{\infty}\frac{1}{m^{2n}}\Delta^n \,. \tag{H.74}$$

As $\Delta$ 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 $\Delta^2$ is:

$$\Delta^2 = \frac{1}{a^4}\begin{pmatrix} 6 & -4 & 1 & & & 1 & -4 \\ -4 & 6 & -4 & 1 & & & \\ 1 & -4 & 6 & -4 & 1 & & \\ & 1 & -4 & \ddots & & & \\ & & & & & 6 & -4 \\ -4 & 1 & & & 1 & -4 & 6 \end{pmatrix} . \tag{H.75}$$

What $\Delta^3$, $\Delta^4$, $\cdots$ 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 H.4.

H.4. **Lattice Laplacian diagonalized.** Insert the identity $\sum \mathbf{P}^{(k)} = \mathbf{1}$ wherever you profitably can, and use the eigenvalue equation (H.50) to convert shift $\mathbf{h}$ matrices into scalars. If $\mathbf{M}$ *commutes* with $\mathbf{h}$, then $(\varphi_k^\dagger \cdot \mathbf{M} \cdot \varphi_{k'}) = \tilde{M}^{(k)}\delta_{kk'}$, and the matrix $\mathbf{M}$ 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 (H.38) the projection on the $k$th subspace is

$$(\varphi_k^\dagger \cdot \Delta \cdot \varphi_{k'}) = \frac{2}{a^2}\left(\cos\left(\frac{2\pi}{N}k\right) - 1\right)\delta_{kk'} \,. \tag{H.76}$$

In the $k$th subspace the propagator is simply a number, and, in contrast to the mess generated by (H.74), there is nothing to evaluating:

$$\varphi_k^\dagger \cdot \frac{1}{m^2\mathbf{1} - \Delta} \cdot \varphi_{k'} = \frac{\delta_{kk'}}{m^2 - \frac{2}{(ma)^2}(\cos 2\pi k/N - 1)} \,, \tag{H.77}$$

where $k$ is a site in the $N$-dimensional dual lattice, and $a = L/N$ is the lattice spacing.

H.5. **Fix Predrag's lecture od Feb 5, 2008.** Are the $C_3$ frequencies on pp. 4,5 correct? If not, write the correct expression for the beat frequency.

## References


[H.1] I. M. Gel'fand, *Lectures on Linear Algebra* (Dover, New York 1961).

[H.2] S. Lang, *Linear Algebra* (Addison-Wesley, Reading, MA 1971).

[H.3] K. Nomizu, *Fundamentals of Linear Algebra* (Chelsea Publ., New York 1979).

# Appendix I

# Convergence of spectral determinants

## I.1 Curvature expansions: geometric picture

If you has some experience with numerical estimates of fractal dimensions, you will note that the numerical convergence of cycle expansions for systems such as the 3-disk game of pinball, table **??**, is very impressive; only three input numbers (the two fixed points $\overline{0}$, $\overline{1}$ and the 2-cycle $\overline{10}$) already yield the escape rate to 4 significant digits! We have omitted an infinity of unstable cycles; so why does approximating the dynamics by a finite number of cycles work so well?

Looking at the cycle expansions simply as sums of unrelated contributions is not specially encouraging: the cycle expansion (18.2) is not absolutely convergent in the sense of Dirichlet series of sect. 18.6, so what one makes of it depends on the way the terms are arranged.

The simplest estimate of the error introduced by approximating smooth flow by periodic orbits is to think of the approximation as a tessalation of a smooth curve by piecewise linear tiles, figure 1.11.

### I.1.1 Tessalation of a smooth flow by cycles

One of the early high accuracy computations of $\pi$ was due to Euler. Euler computed the circumference of the circee of unit radius by inscribing into it a regular polygon with N sides; the error of such computation is proportional to $1 - \cos(2\pi/N) \propto N^{-2}$. In a periodic orbit tessalation of a smooth flow, we cover the phase space by $e^{hn}$ tiles at the $n$th level of resolution, where $h$ is the topological entropy, the growth rate of the number of tiles. Hence we expect the error in approximating a smooth flow by $e^{hn}$ linear segments to be exponentially small, of order $N^{-2} \propto e^{-2hn}$.

### I.1.2 Shadowing and convergence of curvature expansions

We have shown in chapter 13 that if the symbolic dynamics is defined by a finite grammar, a finite number of cycles, let us say the first $k$ terms in the cycle expansion are necessary to correctly count the pieces of the Cantor set generated by the dynamical system.

They are composed of products of non–intersecting loops on the Markov graph, see (13.13). We refer to this set of non–intersecting loops as the *fundamental* cycles of the strange set. It is only after these terms have been included that the cycle expansion is expected to converge smoothly, i.e., only for $n > k$ are the curvatures $c_n$ in (9.2??) a measure of the variation of the quality of a linearized covering of the dynamical Cantor set by the length $n$ cycles, and expected to fall off rapidly with $n$.

The rate of fall-off of the cycle expansion coefficients can be estimated by observing that for subshifts of finite type the contributions from longer orbits in curvature expansions such as (18.7) can always be grouped into shadowing combinations of pseudo-cycles. For example, a cycle with itinerary $\overline{ab}= s_1 s_2 \cdots s_n$ will appear in combination of form

$$1/\zeta = 1 - \cdots - (t_{ab} - t_a t_b) - \cdots ,$$

with $\overline{ab}$ shadowed by cycle $\overline{a}$ followed by cycle $\overline{b}$, where $a = s_1 s_2 \cdots s_m$, $b = s_{m+1} \cdots s_{n-1} s_n$, and $s_k$ labels the Markov partition $\mathcal{M}_{s_k}$ (10.4) that the trajectory traverses at the $k$th return. If the two trajectories coincide in the first $m$ symbols, at the $m$th return to a Poincaré section they can land anywhere in the phase space $\mathcal{M}$

$$\left| f^{T_a}(x_a) - f^{T_{a\cdots}}(x_{a\cdots}) \right| \approx 1 ,$$

where we have assumed that the $\mathcal{M}$ is compact, and that the maximal possible separation across $\mathcal{M}$ is $O(1)$. Here $x_a$ is a point on the $\overline{a}$ cycle of period $T_a$, and $x_{a\cdots}$ is a nearby point whose trajectory tracks the cycle $\overline{a}$ for the first $m$ Poincaré section returns completed at the time $T_{a\cdots}$. An estimate of the maximal separation of the initial points of the two neighboring trajectories is achieved by Taylor expanding around $x_{a\cdots} = x_{\overline{a}} + \delta x_{a\cdots}$

$$f^{T_a}(x_{\overline{a}}) - f^{T_{a\cdots}}(x_{a\cdots}) \approx \frac{\partial f^{T_a}(x_{\overline{a}})}{\partial x} \cdot \delta x_{a\cdots} = M_a \cdot \delta x_{a\cdots} ,$$

hence the hyperbolicity of the flow forces the initial points of neighboring trajectories that track each other for at least $m$ consecutive symbols to lie exponentially close

$$|\delta x_{a\cdots}| \propto \frac{1}{|\Lambda_a|} .$$

Similarly, for any observable (15.1) integrated along the two nearby trajectories

$$A^{T_{a\cdots}}(x_{a\cdots}) \approx A^{T_a}(x_{\overline{a}}) + \left.\frac{\partial A^{T_a}}{\partial x}\right|_{x=x_{\overline{a}}} \cdot \delta x_{a\cdots} ,$$

so

$$\left|A^{T_{a\cdots}}(x_{a\cdots}) - A^{T_a}(x_{\overline{a}})\right| \propto \frac{T_a \mathrm{Const}}{|\Lambda_a|} ,$$

As the time of return is itself an integral along the trajectory, return times of nearby trajectories are exponentially close

$$|T_{a\cdots} - T_a| \propto \frac{T_a \mathrm{Const}}{|\Lambda_a|} ,$$

and so are the trajectory stabilities

$$\left|A^{T_{a\cdots}}(x_{a\cdots}) - A^{T_a}(x_{\overline{a}})\right| \propto \frac{T_a \mathrm{Const}}{|\Lambda_a|} ,$$

Substituting $t_{ab}$ one finds

$$\frac{t_{ab} - t_a t_b}{t_{ab}} = 1 - e^{-s(T_a+T_b-T_{ab})} \left|\frac{\Lambda_a \Lambda_b}{\Lambda_{ab}}\right| .$$

Since with increasing $m$ segments of $\overline{ab}$ come closer to $\overline{a}$, the differences in action and the ratio of the eigenvalues converge exponentially with the eigenvalue of the orbit $\overline{a}$,

$$T_a + T_b - T_{ab} \approx \mathrm{Const} \times \Lambda_a^{-j} , \quad |\Lambda_a \Lambda_b / \Lambda_{ab}| \approx \exp(-\mathrm{Const}/\Lambda_{ab})$$

Expanding the exponentials one thus finds that this term in the cycle expansion is of the order of

$$t_{a^j b} - t_a t_{a^{j-1} b} \approx \mathrm{Const} \times t_{a^j b} \Lambda_a^{-j} . \tag{I.1}$$

Even though the number of terms in a cycle expansion grows exponentially, the shadowing cancellations improve the convergence by an exponential factor compared to trace formulas, and extend the radius of convergence of the periodic orbit sums. Table I.1 shows some examples of such compensations between long cycles and their pseudo-cycle shadows.

| $n$ | $t_{ab} - t_a t_b$ | $T_{ab} - (T_a + T_b)$ | $\log \left\lvert \frac{\Lambda_a \Lambda_b}{\Lambda_{ab}} \right\rvert$ | $ab - a \cdot b$ |
|---|---|---|---|---|
| 2 | -5.23465150784×$10^4$ | 4.85802927371×$10^2$ | -6.3×$10^2$ | 01-0·1 |
| 3 | -7.96028600139×$10^6$ | 5.21713101432×$10^3$ | -9.8×$10^3$ | 001-0·01 |
| 4 | -1.03326529874×$10^7$ | 5.29858199419×$10^4$ | -1.3×$10^3$ | 0001-0·001 |
| 5 | -1.27481522016×$10^9$ | 5.35513574697×$10^5$ | -1.6×$10^4$ | 00001-0·0001 |
| 6 | -1.52544704823×$10^{11}$ | 5.40999882625×$10^6$ | -1.8×$10^5$ | 000001-0·00001 |
| 2 | -5.23465150784×$10^4$ | 4.85802927371×$10^2$ | -6.3×$10^2$ | 01-0·1 |
| 3 | 5.30414752996×$10^6$ | -3.67093656690×$10^3$ | 7.7×$10^3$ | 011-01·1 |
| 4 | -5.40934261680×$10^8$ | 3.14925761316×$10^4$ | -9.2×$10^4$ | 0111-011·1 |
| 5 | 4.99129508833×$10^{10}$ | -2.67292822795×$10^5$ | 1.0×$10^4$ | 01111-0111·1 |
| 6 | -4.39246000586×$10^{12}$ | 2.27087116266×$10^6$ | -1.0×$10^5$ | 011111-01111·1 |

**Table I.1:** Demonstration of shadowing in curvature combinations of cycle weights of form $t_{ab} - t_a t_b$, the 3-disk fundamental domain cycles at $R : d = 6$, table **??**. The ratio $\Lambda_a \Lambda_b / \Lambda_{ab}$ is approaching unity exponentially fast.

It is crucial that the curvature expansion is grouped (and truncated) by topologically related cycles and pseudo-cycles; truncations that ignore topology, such as inclusion of all cycles with $T_p < T_{max}$, will contain orbits unmatched by shadowed orbits, and exhibit a mediocre convergence compared with the curvature expansions.

Note that the existence of a pole at $z = 1/c$ implies that the cycle expansions have a finite radius of convergence, and that analytic continuations will be required for extraction of the non-leading zeros of $1/\zeta$. Preferably, one should work with cycle expansions of Selberg products, as discussed in sect. 18.2.2.

### I.1.3 No shadowing, poorer convergence

Conversely, if the dynamics is not of a finite subshift type, there is no finite topological polynomial, there are no "curvature" corrections, and the convergence of the cycle expansions will be poor.

## I.2 On importance of pruning

If the grammar is not finite and there is no finite topological polynomial, there will be no "curvature" expansions, and the convergence will be poor. That is the generic case, and one strategy for dealing with it is to find a good sequence of approximate but finite grammars; for each approximate grammar cycle expansions yield exponentially accurate eigenvalues, with successive approximate grammars converging toward the desired infinite grammar system.

When the dynamical system's symbolic dynamics does not have a finite grammar, and we are not able to arrange its cycle expansion into curvature combinations (18.7), the series is truncated as in sect. 18.5, by including all pseudo-cycles such that $|\Lambda_{p_1} \cdots \Lambda_{p_k}| \leq |\Lambda_P|$, where $P$ is the most unstable prime cycle included into truncation. The truncation error should then be of order $O(e^{hT_P} T_P / |\Lambda_P|)$, with

$h$ the topological entropy, and $e^{hT_P}$ roughly the number of pseudo-cycles of stability $\approx |\Lambda_P|$. In this case the cycle averaging formulas do not converge significantly better than the approximations such as the trace formula (20.18).

Numerical results (see for example the plots of the accuracy of the cycle expansion truncations for the Hénon map in ref. [3]) indicate that the truncation error of most averages tracks closely the fluctuations due to the irregular growth in the number of cycles. It is not known whether one can exploit the sum rules such as the mass flow conservation (20.11) to improve the accuracy of dynamical averaging.

## I.3 Ma-the-matical caveats

> "Lo duca e io per quel cammino ascoso intrammo a ritornar nel chiaro monde; e sanza cura aver d'alcun riposa salimmo sù, el primo e io secondo, tanto ch'i' vidi de le cose belle che porta 'l ciel, per un perutgio tondo."
>
> —Dante

The periodic orbit theory is learned in stages. At first glance, it seems totally impenetrable. After basic exercises are gone through, it seems totally trivial; all that seems to be at stake are elementary manipulations with traces, determinants, derivatives. But if start thinking about you will get a more and more uncomfortable feeling that from the mathematical point of view, this is a perilous enterprise indeed. In chapter 21 we shall explain which parts of this enterprise are really solid; here you give a fortaste of what objections a mathematician might rise.

Birkhoff's 1931 ergodic theorem states that the time average (15.4) exists almost everywhere, and, if the flow is ergodic, it implies that $\langle a(x)\rangle = \langle a\rangle$ is a constant for almost all $x$. The problem is that the above cycle averaging formulas implicitly rely on ergodic hypothesis: they are strictly correct only if the dynamical system is locally hyperbolic and globally mixing. If one takes a $\beta$ derivative of both sides

$$\rho_\beta(y)e^{ts(\beta)} = \int_{\mathcal{M}} dx\, \delta(y - f^t(x))e^{\beta\cdot A^t(x)}\rho_\beta(x)\,,$$

and integrates over $y$

$$\int_{\mathcal{M}} dy\ \frac{\partial}{\partial\beta}\rho_\beta(y)\Big|_{\beta=0} + t\,\frac{\partial s}{\partial\beta}\Big|_{\beta=0}\int_{\mathcal{M}} dy\,\rho_0(y) = \int_{\mathcal{M}} dx\, A^t(x)\rho_0(x) + \int_{\mathcal{M}} dx\ \frac{\partial}{\partial\beta}\rho_\beta(x)\Big|_{\beta=0}\,,$$

one obtains in the long time limit

$$\left.\frac{\partial s}{\partial \beta}\right|_{\beta=0} = \int_{\mathcal{M}} dy\, \rho_0(x)\, \langle a(x)\rangle \ . \tag{I.2}$$

This is the expectation value (15.12) only if the time average (15.4) equals the space average (15.9), $\langle a(x)\rangle = \langle a\rangle$, for all $x$ except a subset $x \in \mathcal{M}$ of zero measure; if the phase space is foliated into non-communicating subspaces $\mathcal{M} = \mathcal{M}_1 + \mathcal{M}_2$ of finite measure such that $f^t(\mathcal{M}_1) \cap \mathcal{M}_2 = \emptyset$ for all $t$, this fails. In other words, we have tacitly assumed metric indecomposability or transitivity. We have also glossed over the nature of the "phase space" $\mathcal{M}$. For example, if the dynamical system is open, such as the 3-disk game of pinball, $\mathcal{M}$ in the expectation value integral (15.22) is a Cantor set, the closure of the union of all periodic orbits. Alternatively, $\mathcal{M}$ can be considered continuous, but then the measure $\rho_0$ in (I.2) is highly singular. The beauty of the periodic orbit theory is that instead of using an arbitrary coordinatization of $\mathcal{M}$ it partitions the phase space by the intrinsic topology of the dynamical flow and builds the correct measure from cycle invariants, the stability eigenvalues of periodic orbits.

Were we to restrict the applications of the formalism only to systems which have been rigorously proven to be ergodic, we might as well fold up the shop right now. For example, even for something as simple as the Hénon mapping we do not know whether the asymptotic time attractor is strange or periodic. Physics applications require a more pragmatic attitude. In the cycle expansions approach we construct the invariant set of the given dynamical system as a closure of the union of periodic orbits, and investigate how robust are the averages computed on this set. This turns out to depend very much on the observable being averaged over; dynamical averages exhibit "phase transitions", and the above cycle averaging formulas apply in the "hyperbolic phase" where the average is dominated by exponentially many exponentially small contributions, but fail in a phase dominated by few marginally stable orbits. Here the noise - always present, no matter how weak - helps us by erasing an infinity of small traps that the deterministic dynamics might fall into. [exercise 15.1]

Still, in spite of all the caveats, periodic orbit theory is a beautiful theory, and the cycle averaging formulas are the most elegant and powerful tool available today for evaluation of dynamical averages for low dimensional chaotic deterministic systems.

## I.4 Estimate of the *n*th cumulant

An immediate consequence of the exponential spacing of the eigenvalues is that the convergence of the Selberg product expansion (D.12) as function of the topological cycle length, $F(z) = \sum_n C_n z^n$, is faster than exponential. Consider a $d$–dimensional map for which all fundamental matrix eigenvalues are equal: $u_p = \Lambda_{p,1} = \Lambda_{p,2} = \cdots = \Lambda_{p,d}$. The stability eigenvalues are generally not isotropic; however, to obtain qualitative bounds on the spectrum, we replace all

stability eigenvalues with the least expanding one. In this case the $p$ cycle contribution to the product (17.9) reduces to

$$
\begin{aligned}
F_p(z) &= \prod_{k_1\cdots k_d=0}^{\infty}\left(1-t_p u_p^{k_1+k_2+\cdots+k_d}\right) \\
&= \prod_{k=0}^{\infty}\left(1-t_p u_p^{k}\right)^{m_k}\,; \qquad m_k=\binom{d-1+k}{d-1}=\frac{(k+d-1)!}{k!(d-1)!} \\
&= \prod_{k=0}^{\infty}\sum_{\ell=0}^{m_k}\binom{m_k}{\ell}\left(-u_p^k t_p\right)^{\ell}
\end{aligned}
\tag{I.3}
$$

In one dimension the expansion can be given in closed form (21.34), and the coefficients $C_k$ in (D.12) are given by

$$
\tau_{p^k}=(-1)^k\frac{u_p^{\frac{k(k-1)}{2}}}{\prod_{j=1}^{k}(1-u_p^j)}t_p^k\ . \tag{I.4}
$$

Hence the coefficients in the $F(z)=\sum_n C_n z^n$ expansion of the spectral determinant (18.11) fall off faster than exponentially, as $|C_n|\approx u^{n(n-1)/2}$. In contrast, the cycle expansions of dynamical zeta functions fall of "only" exponentially; in numerical applications, the difference is dramatic.

In higher dimensions the expansions are not quite as compact. The leading power of $u$ and its coefficient are easily evaluated by use of binomial expansions (I.3) of the $(1+tu^k)^{m_k}$ factors. More precisely, the leading $u^n$ terms in $t^k$ coefficients are of form

$$
\begin{aligned}
\prod_{k=0}^{\infty}(1+tu^k)^{m_k} &= \ldots+u^{m_1+2m_2+\ldots+jm_j}\,t^{1+m_1+m_2+\ldots+m_j}+\ldots \\
&= \ldots+\left(u^{\frac{m_d}{d+1}}t\right)^{\binom{d+m}{m}}+\ldots\approx\ldots+u^{\frac{\sqrt[d]{d!}}{(d-1)!}n^{\frac{d+1}{d}}}t^n+\ldots
\end{aligned}
$$

Hence the coefficients in the $F(z)$ expansion fall off faster than exponentially, as $u^{n^{1+1/d}}$. The Selberg products are entire functions in any dimension, provided that the symbolic dynamics is a finite subshift, and all cycle eigenvalues are sufficiently bounded away from 1.

The case of particular interest in many applications are the 2-d Hamiltonian mappings; their symplectic structure implies that $u_p=\Lambda_{p,1}=1/\Lambda_{p,2}$, and the Selberg product (17.13) In this case the expansion corresponding to (21.34) is given by (21.35) and the coefficients fall off asymptotically as $C_n\approx u^{n^{3/2}}$.

# Appendix J

# Infinite dimensional operators

(A. Wirzba)

THIS APPENDIX, taken from ref. [1], 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.

## J.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

$$A'(x) = \frac{dA(x)}{dx}\,, \qquad A'_{ij}(x) = \frac{d}{dx}A_{ij}(x)\,. \tag{J.1}$$

Derivatives of products of matrices are evaluated by the chain rule

$$\frac{d}{dx}(A\mathbf{B}) = \frac{dA}{dx}\mathbf{B} + A\frac{d\mathbf{B}}{dx}\,. \tag{J.2}$$

A matrix and its derivative matrix in general do not commute

$$\frac{d}{dx}A^2 = \frac{dA}{dx}A + A\frac{dA}{dx}\,. \tag{J.3}$$

The derivative of the inverse of a matrix, follows from $\frac{d}{dx}(AA^{-1}) = 0$:

$$\frac{d}{dx}A^{-1} = -\frac{1}{A}\frac{dA}{dx}\frac{1}{A}\,. \tag{J.4}$$

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:

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k , \qquad A^0 = \mathbf{1} \tag{J.5}$$

$$= \lim_{N\to\infty} \left( \mathbf{1} + \frac{1}{N} A \right)^N \tag{J.6}$$

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 $\lim \delta x_N \to 0$ as $N \to \infty$, the extra terms do not contribute.

Consider now the determinant

$$\det(e^A) = \lim_{N\to\infty} (\det\,(\mathbf{1} + A/N))^N .$$

To the leading order in $1/N$

$$\det\,(\mathbf{1} + A/N) = 1 + \frac{1}{N} \mathrm{tr}\, A + O(N^{-2}) .$$

hence

$$\det e^A = \lim_{N\to\infty} \left(1 + \frac{1}{N} \mathrm{tr}\, A + O(N^{-2})\right)^N = e^{\mathrm{tr}\, A} \tag{J.7}$$

Due to non-commutativity of matrices, generalization of a function of several variables to a function is not as straightforward. Expression involving several matrices depend on their commutation relations. For example, the commutator expansion

$$e^{t\mathbf{A}} \mathbf{B} e^{-t\mathbf{A}} = \mathbf{B} + t[\mathbf{A}, \mathbf{B}] + \frac{t^2}{2}[\mathbf{A}, [\mathbf{A}, \mathbf{B}]] + \frac{t^3}{3!}[\mathbf{A}, [\mathbf{A}, [\mathbf{A}, \mathbf{B}]]] + \cdots \tag{J.8}$$

sometimes used to establish the equivalence of the Heisenberg and Schrödinger pictures of quantum mechanics follows by recursive evaluation of $t$ derivatives

$$\frac{d}{dt}\left(e^{t\mathbf{A}} \mathbf{B} e^{-t\mathbf{A}}\right) = e^{t\mathbf{A}} [\mathbf{A}, \mathbf{B}] e^{-t\mathbf{A}} .$$

Manipulations of such ilk yield

$$e^{(\mathbf{A}+\mathbf{B})/N} = e^{\mathbf{A}/N} e^{\mathbf{B}/N} - \frac{1}{2N^2}[\mathbf{A},\mathbf{B}] + O(N^{-3}) \,,$$

and the Trotter product formula: if $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{A} = \mathbf{B} + \mathbf{C}$ are matrices, then

$$e^{\mathbf{A}} = \lim_{N\to\infty} \left(e^{\mathbf{B}/N} e^{\mathbf{C}/N}\right)^N \tag{J.9}$$

## J.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 $n \to \infty$ convergence of partial products

$$\mathbf{E}_n = \prod_{0\le m<n} \left(\mathbf{1} + \frac{t}{m} A\right)$$

can be verified using the Cauchy criterion, which states that the sequence $\{\mathbf{E}_n\}$ converges if the differences $\|\mathbf{E}_k - \mathbf{E}_j\| \to 0$ as $k, j \to \infty$. To make sense of this we need to define a sensible norm $\|\cdots\|$. 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 $\hat{n}$:

$$\|\mathbf{M}\| \;=\; \sup_{\hat{n}} \|\mathbf{M}\hat{n}\| \,, \qquad \|\hat{n}\| = 1 \,. \tag{J.10}$$

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

$$\|\mathbf{M}\| = \max_i \sum_j |M_{ij}| \,. \tag{J.11}$$

The operator norm (J.11) and the vector norm (J.10) 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

[exercise J.1]

$$||e^{tA}|| \leq e^{t||A||} \,. \tag{J.12}$$

[exercise 2.9]

As $||A||$ is a number, the norm of $e^{tA}$ 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 (4.10) has a solution for all times.

## J.3 Trace class and Hilbert-Schmidt class

This section is mainly an extract from ref. [9]. Refs. [7, 10, 11, 14] 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 $\mathbf{A} \in \mathcal{L}(\mathcal{H})$: $\mathcal{H} \to \mathcal{H}$ (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. J.5 where the theory of *von Koch matrices* of ref. [12] 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, $\{\phi_n\}$:

$$\sum_n |\langle \phi_n, \mathbf{A}\phi_n \rangle| < \infty \,. \tag{J.13}$$

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 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 $||B|| = \sup_{\phi \neq 0} \frac{||\mathbf{B}\phi||}{||\phi||}$ 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 $\langle \mathbf{A}\phi, \phi \rangle \geq 0 \ \forall \phi \in \mathcal{H}$. Notice 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, 9]):

**(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 $\mathbf{A} = \mathbf{BC}$ with $\mathbf{B}, \mathbf{C} \in \mathcal{J}_2$.

**(c)** $\mathcal{J}_1 \subset \mathcal{J}_2 \subset$ 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 $\mathbf{A} \in \mathcal{J}_1$, $\mathrm{Tr}(\mathbf{A}) = \sum \langle \phi_n, \mathbf{A}\phi_n \rangle$ is independent of the basis used.

**(f)** Tr is linear and obeys $\mathrm{Tr}(\mathbf{A}^\dagger) = \overline{\mathrm{Tr}(\mathbf{A})}$; $\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})$ 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 $\langle \mathbf{A}, \mathbf{B} \rangle_2 = \mathrm{Tr}(\mathbf{A}^\dagger \mathbf{B})$ is a Hilbert space. If $\|\mathbf{A}\|_2 = [\,\mathrm{Tr}(\mathbf{A}^\dagger \mathbf{A})\,]^{\frac{1}{2}}$, then $\|\mathbf{A}\|_2 \geq \|\mathbf{A}\|$ and $\mathcal{J}_2$ is the $\| \ \|_2$-closure of the *finite* rank operators.

**(h)** $\mathcal{J}_1$ endowed with the norm $\|\mathbf{A}\|_1 = \mathrm{Tr}(\sqrt{\mathbf{A}^\dagger \mathbf{A}})$ is a Banach space. $\|\mathbf{A}\|_1 \geq \|\mathbf{A}\|_2 \geq \|\mathbf{A}\|$ and $\mathcal{J}_1$ is the $\| \ \|_1$-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 $\langle \mathbf{B}, \mathbf{A} \rangle = \mathrm{Tr}(\mathbf{BA})$.

**(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 $\mathbf{B}$.

## J.4 Determinants of trace class operators

This section is mainly based on refs. [8, 10] which should be consulted for more details and proofs. See also refs. [11, 14].

**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 $\phi_1, \ldots, \phi_n \in \mathcal{H}$. $\bigwedge^n(\mathcal{H})$ 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 $\pi$ is an even or odd permutation, respectively. The inner product in $\bigwedge^n(\mathcal{H})$ is given by

$$(\phi_1 \wedge \cdots \wedge \phi_n, \eta_1 \wedge \cdots \wedge \eta_n) = \det\Big\{(\phi_i, \eta_j)\Big\}$$

where $\det\{a_{ij}\} = \sum_{\pi\in\mathcal{P}_n} \epsilon(\pi) a_{1\pi(1)} \cdots a_{n\pi(n)}$. $\bigwedge^n(\mathbf{A})$ is defined as functor (a functor satisfies $\bigwedge^n(\mathbf{AB}) = \bigwedge^n(\mathbf{A}) \bigwedge^n(\mathbf{B})$) on $\bigwedge^n(\mathcal{H})$ with

$$\bigwedge^n(\mathbf{A})\,(\phi_1 \wedge \cdots \wedge \phi_n) = \mathbf{A}\phi_1 \wedge \cdots \wedge \mathbf{A}\phi_n\,.$$

When $n = 0$, $\bigwedge^n(\mathcal{H})$ is defined to be $C$ and $\bigwedge^n(\mathbf{A})$ as $1\colon C \to C$.

**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

$$\mathrm{Tr}\Big(\bigwedge^k(\mathbf{A})\Big) = \sum_{i_1<\cdots<i_k} ((\phi_{i_1} \wedge \cdots \wedge \phi_{i_k}), (\mathbf{A}\phi_{i_1} \wedge \cdots \wedge \mathbf{A}\phi_{i_k})) < \infty$$

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

$$\det(\mathbf{1} + \mathbf{A}) = \sum_{k=0}^{\infty} \mathrm{Tr}\Big(\bigwedge^k(\mathbf{A})\Big) \tag{J.14}$$

**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)** $|\det(\mathbf{1}+\mathbf{A})| \leq \prod_{j=1}^{\infty}\left(1+\mu_j(\mathbf{A})\right)$ 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 $\mathbf{A}_1, \ldots, \mathbf{A}_n \in \mathcal{J}_1$, $\langle z_1, \ldots, z_n\rangle \mapsto \det\left(\mathbf{1}+\sum_{i=1}^n z_i\mathbf{A}_i\right)$ is an entire analytic function.

**(e)** If $\mathbf{A}, \mathbf{B} \in \mathcal{J}_1$, then

$$\begin{aligned}\det(\mathbf{1}+\mathbf{A})\det(\mathbf{1}+\mathbf{B}) &= \det(\mathbf{1}+\mathbf{A}+\mathbf{B}+\mathbf{AB}) \\ &= \det((\mathbf{1}+\mathbf{A})(\mathbf{1}+\mathbf{B})) \\ &= \det((\mathbf{1}+\mathbf{B})(\mathbf{1}+\mathbf{A}))\ . \end{aligned} \tag{J.15}$$

If $\mathbf{A} \in \mathcal{J}_1$ and $\mathbf{U}$ unitary, then

$$\det\left(\mathbf{U}^{-1}(\mathbf{1}+\mathbf{A})\mathbf{U}\right) = \det\left(\mathbf{1}+\mathbf{U}^{-1}\mathbf{AU}\right) = \det(\mathbf{1}+\mathbf{A})\ .$$

**(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 $\epsilon$, 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 $\mathbf{A} \in \mathcal{J}_1$,

$$\det(\mathbf{1}+\mathbf{A}) = \prod_{j=1}^{N(\mathbf{A})}\left(1+\lambda_j(\mathbf{A})\right) \tag{J.16}$$

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 $\mathbf{A} \in \mathcal{J}_1$,

$$\mathrm{Tr}(\mathbf{A}) = \sum_{j=1}^{N(\mathbf{A})} \lambda_j(\mathbf{A}) < \infty\ .$$

**(k)** If $\mathbf{A} \in \mathcal{J}_1$, then

$$\begin{aligned}\mathrm{Tr}\left(\bigwedge^k(\mathbf{A})\right) &= \sum_{j=1}^{N\left(\bigwedge^k(\mathbf{A})\right)} \lambda_j\left(\bigwedge^k(\mathbf{A})\right) \\ &= \sum_{1\leq j_1<\cdots<j_k\leq N(\mathbf{A})} \lambda_{j_1}(\mathbf{A})\cdots\lambda_{j_k}(\mathbf{A}) < \infty.\end{aligned}$$

**(l)** If $\mathbf{A} \in \mathcal{J}_1$, then

$$\det(1 + z\mathbf{A}) = \sum_{k=0}^{\infty} z^k \sum_{1 \le j_1 < \cdots < j_k \le N(\mathbf{A})} \lambda_{j_1}(\mathbf{A}) \cdots \lambda_{j_k}(\mathbf{A}) < \infty. \tag{J.17}$$

**(m)** If $\mathbf{A} \in \mathcal{J}_1$, then for $|z|$ small (i.e., $|z| \max|\lambda_j(\mathbf{A})| < 1$) the series $\sum_{k=1}^{\infty} z^k \mathrm{Tr}\left((-\mathbf{A})^k\right)/k$ converges and

$$\begin{aligned} \det(1 + z\mathbf{A}) &= \exp\left(-\sum_{k=1}^{\infty} \frac{z^k}{k} \mathrm{Tr}\left((-\mathbf{A})^k\right)\right) \\ &= \exp\left(\mathrm{Tr}\ln(\mathbf{1} + z\mathbf{A})\right) . \end{aligned} \tag{J.18}$$

**(n)** *The Plemelj-Smithies formula:* Define $\alpha_m(\mathbf{A})$ for $\mathbf{A} \in \mathcal{J}_1$ by

$$\det(\mathbf{1} + z\mathbf{A}) = \sum_{m=0}^{\infty} z^m \frac{\alpha_m(\mathbf{A})}{m!} . \tag{J.19}$$

Then $\alpha_m(\mathbf{A})$ is given by the $m \times m$ determinant:

$$\alpha_m(\mathbf{A}) = \begin{vmatrix} \mathrm{Tr}(\mathbf{A}) & m-1 & 0 & \cdots & 0 \\ \mathrm{Tr}(\mathbf{A}^2) & \mathrm{Tr}(\mathbf{A}) & m-2 & \cdots & 0 \\ \mathrm{Tr}(\mathbf{A}^3) & \mathrm{Tr}(\mathbf{A}^2) & \mathrm{Tr}(\mathbf{A}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & 1 \\ \mathrm{Tr}(\mathbf{A}^m) & \mathrm{Tr}(\mathbf{A}^{(m-1)}) & \mathrm{Tr}(\mathbf{A}^{(m-2)}) & \cdots & \mathrm{Tr}(\mathbf{A}) \end{vmatrix} \tag{J.20}$$

with the understanding that $\alpha_0(\mathbf{A}) \equiv 1$ and $\alpha_1(\mathbf{A}) \equiv \mathrm{Tr}(\mathbf{A})$. Thus the cumulants $c_m(\mathbf{A}) \equiv \alpha_m(\mathbf{A})/m!$ satisfy the following recursion relation

$$\begin{aligned} c_m(\mathbf{A}) &= \frac{1}{m} \sum_{k=1}^{m} (-1)^{k+1} c_{m-k}(\mathbf{A}) \mathrm{Tr}(\mathbf{A}^k) \quad \text{for } m \ge 1 \\ c_0(\mathbf{A}) &\equiv 1 . \end{aligned} \tag{J.21}$$

Note that in the context of quantum mechanics formula (J.19) is the quantum analog to the curvature expansion of the semiclassical zeta function with $\mathrm{Tr}(\mathbf{A}^m)$ corresponding to the sum of all periodic orbits (prime and also repeated ones) of *total* topological length $m$, i.e., let $c_m(\mathrm{s.c.})$ denote the $m^{\mathrm{th}}$ curvature term, then the curvature expansion of the semiclassical zeta function is given by the recursion relation

$$\begin{aligned} c_m(\mathrm{s.c.}) &= \frac{1}{m} \sum_{k=1}^{m} (-1)^{k+m+1} c_{m-k}(\mathrm{s.c.}) \sum_{\substack{p;r>0 \\ \text{with } [p]r=k}} [p] \frac{t_p(k)^r}{1 - \left(\frac{1}{\Lambda_p}\right)^r} \quad \text{for } m \ge 1 \\ c_0(\mathrm{s.c.}) &\equiv 1 . \end{aligned} \tag{J.22}$$

In fact, in the cumulant expansion (J.19) as well as in the curvature expansion there are large cancelations involved. Let us order – without lost of generality – the eigenvalues of the operator $\mathbf{A} \in \mathcal{J}_1$ as follows:

$$|\lambda_1| \ge |\lambda_2| \ge \cdots \ge |\lambda_{i-1}| \ge |\lambda_i| \ge |\lambda_{i+1}| \ge \cdots$$

(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. (J.19), we have enormous cancelations of big numbers, e.g. at the $k^{\text{th}}$ cumulant order ($k > 3$), all the intrinsically large 'numbers' $\lambda_1^k$, $\lambda_1^{k-1}\lambda_2$, ..., $\lambda_1^{k-2}\lambda_2\lambda_3$, ... and many more have to cancel out exactly until only $\sum_{1\le j_1<\cdots<j_k\le N(\mathbf{A})} \lambda_{j_1}\cdots\lambda_{j_k}$ 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 (J.17) or even to (J.16) directly. It might not be possible to find a direct semiclassical analog for the individual eigenvalues $\lambda_j$. Thus the direct construction of the semiclassical equivalent to (J.16) is rather unlikely. However, in order to have a semiclassical "cumulant" summation without large cancelations – see (J.17) – it would be just sufficient to find the semiclassical analog of each complete cumulant (J.17) and not of the single eigenvalues. Whether this will eventually be possible is still an open question.

## J.5 Von Koch matrices

Implicitly, many of the above properties are based on the theory of von Koch matrices [11, 12, 13]: 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 $\sum |a_{j_1k_1}a_{j_2k_2}\cdots a_{j_n,k_n}|$ converges, where the sum extends over all pairs of systems of indices $(j_1, j_2, \cdots, j_n)$ and $(k_1, k_2, \cdots, k_n)$ which differ from each other only by a permutation, and $j_1 < j_2 < \cdots j_n$ ($n = 1, 2, \cdots$). Then the limit

$$\lim_{n\to\infty} \det\|\delta_{jk} - a_{jk}\|_1^n = \det(\mathbf{1} - \mathbf{A})$$

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 $a_{jk}$ ($j, k = 1, 2, \cdots$) 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

$$\sum_{j=1}^{\infty} |a_{jj}| \quad < \quad \infty \,, \tag{J.23}$$

$$\sum_{j,k=1}^{\infty} |a_{jk}|^2 \quad < \quad \infty \tag{J.24}$$

are fulfilled. Then the following holds (see ref. [11, 13]): (1) Every von Koch matrix has an absolutely convergent determinant. If the elements of a von Koch matrix are functions of some parameter $\mu$ ($a_{jk} = a_{jk}(\mu)$, $j,k = 1,2,\cdots$) and both series in the defining condition converge uniformly in the domain of the parameter $\mu$, then as $n \to \infty$ the determinant $\det\|\delta_{jk} - a_{jk}(\mu)\|_1^n$ tends to the determinant $\det(\mathbf{1}+\mathbf{A}(\mu))$ uniformly with respect to $\mu$, over the domain of $\mu$. (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 (J.24) 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:

$$\mathbf{A}_{ij} = \begin{cases} 2/j & \text{for} \quad i-j=-1 \quad \text{and} \quad j \text{ even}\,, \\ 2/i & \text{for} \quad i-j=+1 \quad \text{and} \quad i \text{ even}\,, \\ 0 & \text{else}\,, \end{cases}$$

i.e.,

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1/2 & 0 & 0 & \cdots \\ 0 & 0 & 1/2 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 1/3 & \ddots \\ 0 & 0 & 0 & 0 & 1/3 & 0 & \ddots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}. \tag{J.25}$$

Obviously, condition (J.23) is fulfilled by definition. Second, the condition (J.24) 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 (J.25) violates the trace-class definition (J.13), 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).

## J.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 $\mathrm{Tr}|A|^p < \infty$ 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 [8, 10]:
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

$$\det_n(\mathbf{1} + z\mathbf{A}) = \prod_{j=1}^{N(z\mathbf{A})} \left[\left(1 + z\lambda_j(\mathbf{A})\right) \exp\left(\sum_{k=1}^{n-1} \frac{\left(-z\lambda_j(\mathbf{A})\right)^k}{k}\right)\right] < \infty. \tag{J.26}$$

exists and is finite. The corresponding Plemelj-Smithies formula now reads [10]:

$$\det_n(\mathbf{1} + z\mathbf{A}) = \sum_{m=0}^{\infty} z^m \frac{\alpha_m^{(n)}(\mathbf{A})}{m!} \,. \tag{J.27}$$

with $\alpha_m^{(n)}(\mathbf{A})$ given by the $m \times m$ determinant:

$$\alpha_m^{(n)}(\mathbf{A}) = \begin{vmatrix} \sigma_1^{(n)} & m-1 & 0 & \cdots & 0 \\ \sigma_2^{(n)} & \sigma_1^{(n)} & m-2 & \cdots & 0 \\ \sigma_3^{(n)} & \sigma_2^{(n)} & \sigma_1^{(n)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & 1 \\ \sigma_m^{(n)} & \sigma_{m-1}^{(n)} & \sigma_{m-2}^{(n)} & \cdots & \sigma_1^{(n)} \end{vmatrix} \tag{J.28}$$

where

$$\sigma_k^{(n)} = \begin{cases} \mathrm{Tr}(\mathbf{A}^k) & k \geq n \\ 0 & k \leq n-1 \end{cases}$$

As Simon [10] says simply, the beauty of (J.28) is that we get $\det_n(\mathbf{1} + \mathbf{A})$ from the standard Plemelj-Smithies formula (J.19) by simply setting $\mathrm{Tr}(\mathbf{A})$, $\mathrm{Tr}(\mathbf{A}^2)$, ..., $\mathrm{Tr}(\mathbf{A}^{n-1})$ to zero.

See also ref. [15] 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 $d$, $0 < \lambda_0 \leq \lambda_1 \leq \cdots$ and $\lambda_k \uparrow +\infty$. and the Fredholm determinant

$$\Delta(\lambda) = \prod_{k=0}^{\infty} \left(1 - \frac{\lambda}{\lambda_k}\right)$$

is regulated in the case $\mu \equiv d/m > 1$ as Weierstrass product

$$\Delta(\lambda) = \prod_{k=0}^{\infty} \left[\left(1 - \frac{\lambda}{\lambda_k}\right) \exp\left(\frac{\lambda}{\lambda_k} + \frac{\lambda^2}{2\lambda_k^2} + \cdots + \frac{\lambda^{[\mu]}}{[\mu]\lambda_k^{[\mu]}}\right)\right] \tag{J.29}$$

where $[\mu]$ denotes the integer part of $\mu$. This is, see ref. [15], the unique entire function of order $\mu$ 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 (J.29) is the same as (J.26); 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

$$\Delta(E) = \det\left[(E - \mathbf{H})\right] \tag{J.30}$$

which – as it stands – would only make sense for a finite dimensional basis (or finite dimensional matrices). In ref. [16] 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

$$\Delta(E) = \det\left[(E - \mathbf{H})\Omega(E, \mathbf{H})\right]$$

where

$$\Omega(E, \mathbf{H}) = -\mathbf{H}^{-1} e^{E\mathbf{H}^{-1}}$$

such that the spectral determinant in the eigenbasis of **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.

## Exercises

J.1. **Norm of exponential of an operator.** Verify inequality (J.12):

$$\|e^{tA}\| \leq e^{t\|A\|} \ .$$

## References

[J.1] A. Wirzba, *Quantum Mechanics and Semiclassics of Hyperbolic n-Disk Scattering*, Habilitationsschrift, Technische Universität, Germany, 1997, HAB, chao-dyn/9712015, *Physics Reports* in press.

[J.2] A. Grothendieck, *"La théorie de Fredholm," Bull. Soc. Math. France*, **84**, 319 (1956).

[J.3] A. Grothendieck, *Produits tensoriels topologiques et espaces nucléaires*, Amer. Meth. Soc. **16**, Providence R. I. (1955).

[J.4] C.A. Tracy and H. Widom, CHECK THIS!: *Fredholm Determinants, Differential Equations and Matrix Models*, hep-th/9306042.

[J.5] M.G. Krein, *On the Trace Formula in Perturbation Theory* Mat.. Sborn. (N.S.) **33** (1953) 597-626; *Perturbation Determinants and Formula for Traces of Unitary and Self-adjoint Operators* Sov. Math.-Dokl. **3** (1962) 707-710. M.S. Birman and M.G. Krein, *On the Theory of Wave Operators and Scattering Operators*, Sov. Math.-Dokl. **3** (1962) 740-744.

[J.6] J. Friedel, Nuovo Cim. Suppl. **7** (1958) 287-301.

[J.7] M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. I: *Functional Analysis*, Chap. VI, Academic Press (New York), 1972.

[J.8] M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. IV: *Analysis of Operators*, Chap. XIII.17, Academic Press (New York), 1976.

[J.9] B. Simon, *Quantum Mechanics for Hamiltonians defined as Quadratic Forms*, Princeton Series in Physics, 1971, Appendix.

[J.10] B. Simon, *Notes on Infinite Determinants of Hilbert Space Operators*, Adv. Math. **24** (1977) 244-273.

[J.11] I.C. Gohberg and M.G. Krein, *Introduction to the theory of linear non-selfadjoint operators*, Translations of Mathematical Monographs **18**, Amer. Math. Soc. (1969).

[J.12] H. von Koch, *Sur quelques points de la théorie des déterminants infinis*, Acta. Math. **24** (1900) 89-122; *Sur la convergence des déterminants infinis*, Rend. Circ. Mat. Palermo **28** (1909) 255-266.

[J.13] E. Hille and J.D. Tamarkin, *On the characteristic values of linear integral equations*, Acta Math. **57** (1931) 1-76.

[J.14] T. Kato, *Perturbation Theory of Linear Operators* (Springer, New York, 1966), Chap. X, § 1.3 and § 1.4.

[J.15] A. Voros, *Spectral Functions, Special Functions and the Selberg Zeta Function*, *Comm. Math Phys.* **110**, 439 (1987).

[J.16] J.P. Keating and M. Sieber, "Calculation of spectral determinants," preprint (1994).

# Appendix K

# 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.

## K.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 interst 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 Boltzman 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 $\beta$, 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.

$$f(\beta) \to \frac{1}{n} \tag{K.1}$$

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,

$$f(\beta) \to e^{-\alpha n} \tag{K.2}$$

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 $\alpha$ 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 $\alpha$ 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 $Z_n$

$$Z_n = \mathrm{tr}\ T^n$$

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 described here solves both of these problems in a limited context: it regularizes the transfer operator in a physically meaningful way, and as a 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.

## K.2 Ising models

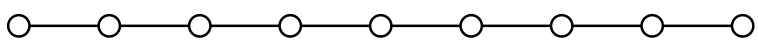
The Ising model is a simple model to study the cooperative effects of many small interacting magnetic dipoles. The dipoles are placed on a lattice and their interaction is greatly simplified. There can also be a field that includes the effects of an external magnetic field and the average effect of the 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.

### K.2.1 Ising model

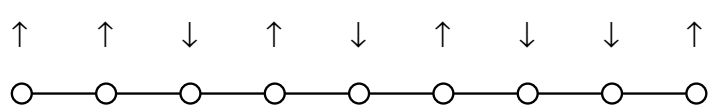
One of the simplest models in statistical mechanics is the Ising model. One imagines that one has a 1-dimensional lattice with small magnets at each site that can point either up or down.

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 chsen 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



(K.3)

(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 (↑) or down (↓). This represents the two states of the small magnet on that site, and in general we will use the notation $\Sigma_0$ 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



(K.4)

The set of all configurations for a lattice with $n$ sites is called $\Omega_0^n$ 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

$$\sigma = \{\sigma_0, \sigma_1, \ldots \sigma_n\}, \tag{K.5}$$

In the example configuration (K.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

$$H(\sigma) = \sum_i J\delta(\sigma_i, \sigma_{i+1}), \tag{K.6}$$

where

$$\delta(\sigma_1, \sigma_2) = \begin{cases} +1 & \text{if } \sigma_1 = \sigma_2 \\ -1 & \text{if } \sigma_1 \neq \sigma_2 \end{cases} . \tag{K.7}$$

One of the problems that was avoided when computing the energy was what to do at the boundaries of the 1-dimensional chain. Notice that as written, (K.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 $\beta$ is given by

$$-\beta f(\beta) = \lim_{n\to\infty} \frac{1}{n} \ln Z^{(n)} . \tag{K.8}$$

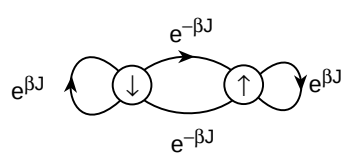
The partition function $Z^{(n)}$ is computed by a sum that runs over all the possible configurations on the 1-dimensional chain. Each configuration contributes with its Gibbs factor $\exp(-\beta H(\sigma))$ and the partition function $Z^{(n)}$ is

$$Z^{(n)}(\beta) = \sum_{\sigma\in\Omega_0^n} e^{-\beta H(\sigma)} . \tag{K.9}$$

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

$$Z^{(n)}(\beta) = \sum_{\sigma\in\Omega^n} e^{\beta J\delta(\sigma_1,\sigma_2)} e^{\beta J\delta(\sigma_2,\sigma_3)} \cdots e^{\beta J\delta(\sigma_n,\sigma_1)} . \tag{K.10}$$

Let us use the configuration in (K.4). The first site is $\sigma_1 = \uparrow$. As the second site is $\uparrow$, we know that the first term in (K.10) is a term $e^{\beta J}$. The third spin is $\downarrow$, so the second term in (K.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 $\sigma_1$. 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 Markov 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 Markov 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

$$T(\beta) = \begin{bmatrix} e^{\beta J} & e^{-\beta J} \\ e^{-\beta J} & e^{\beta J} \end{bmatrix} . \tag{K.11}$$

As in chapter 10 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 $\lambda_1$ and $\lambda_2$ of the transfer matrix:

$$Z^{(n)}(\beta) = \operatorname{tr} T^n(\beta) = \lambda_1(\beta)^n + \lambda_2(\beta)^n . \tag{K.12}$$

### K.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?

### K.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 assumes a finite number of states identified by the set $\Omega_0$.

The transfer operator $\mathcal{T}$ was introduced by Kramers and Wannier [12] 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.

## K.3 Fisher droplet model

In a series of articles [20], 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 [20]. Gallavotti [21] 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 [22]. 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.

### K.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 $\theta_n$ 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 $\theta_1$. 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 $\theta_n$ term to the energy. The total energy of the configuration is then

$$H(0111010) = 4\theta_1 + 2\theta_2 + 1\theta_3 \,. \tag{K.13}$$

If there where 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

$$a = \sum_{n>0} \theta_n < \infty \,. \tag{K.14}$$

The configuration with all zeros does not contribute to the energy.

Once we specify the function $\theta_n$ we can computed 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

$$G(z,\beta) = \sum_{n>0} z^n \frac{Z_n(\beta)}{n} \tag{K.15}$$

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

$$\ldots 011011011011 \ldots$$

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

$$G(z,\beta) = \sum_{b\in B} \frac{z^{|b|}}{s(b)} e^{-\beta H(b)} \,. \tag{K.16}$$

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 |
| 3 | 001 011 |
| 4 | 0001 0011 0111 |

(K.17)

Each droplet of size $n$ contributes with energy

$$W_n = \sum_{1\le k\le n} (n-k+1)\theta_k \,. \tag{K.18}$$

So if we consider the sum

$$\sum_{n\ge 1} \frac{1}{n}\Big(z^2 e^{-\beta H(01)} + z^3(e^{-\beta H(001)} + e^{-\beta H(011)}) + \\ + z^4(e^{-\beta H(0001)} + e^{-\beta H(0011)} + e^{-\beta H(0111)}) + \cdots\Big)^n \tag{K.19}$$

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

$$-\ln\Big(1 - (z^2 e^{-\beta W_1} + z^3(e^{-\beta W_1} + e^{-\beta W_2}) + \cdots\Big) \,, \tag{K.20}$$

where the $H(b)$ factors have been evaluated in terms of the droplet energies $W_n$. A proof of the equality of (K.19) and (K.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

$$-\ln\Big(1 - (z^1 + z^2 + z^3 + \cdots)(ze^{-\beta W_1} + z^2 e^{-\beta W_2} + \cdots)\Big) \tag{K.21}$$

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

$$\sum_{n>0} \frac{z^n}{n} \,. \tag{K.22}$$

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

$$\sum_{n>0} \frac{z^n e^{-\beta H(\overbrace{11\ldots11}^{n})}}{n} . \tag{K.23}$$

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 (K.21), (K.22), and (K.23) we can evaluate the generating function $G$ to be

$$G(z,\beta) = -\ln(1-z) - \ln(1-ze^{-\beta a}) - \ln(1 - \frac{z}{1-z}\sum_{n\geq 1} z^n e^{-\beta W_n}) . \tag{K.24}$$

The cycle expansion $\zeta^{-1}(z,\beta)$ is given by the exponential of the generating function $e^{-G}$ and we obtain

$$\zeta^{-1}(z,\beta) = (1-ze^{-\beta a})(1 - z(1 + \sum_{n\geq 1} z^n e^{-\beta W_n})) \tag{K.25}$$

To pursue this model further we need to have some assumptions about the interaction strengths $\theta_n$. 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

$$W_n \sim -n + \ln n + O(\frac{1}{n}) . \tag{K.26}$$

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

$$W_n = -n + \ln n \tag{K.27}$$

in a system of units where the dimensional constants are one. To evaluate the cycle expansion (K.25) we need to evaluate the constant $a$, the sum of all the $\theta_n$. One can write a recursion for the $\theta_n$

$$\theta_n = W_n - \sum_{1\leq k<n} (n-k+1)\theta_k \tag{K.28}$$

and with an initial choice for $\theta_1$ evaluate all the others. It can be verified that independent of the choice of $\theta_1$ the constant $a$ is equal to the number that multiplies the $n$ term in (K.27). In the units used

$$a = -1\,. \tag{K.29}$$

For the choice of droplet energy (K.27) the sum in the cycle expansion can be expressed in terms of a special function: the Lerch transcendental $\phi_L$. It is defined by

$$\phi_L(z, s, c) = \sum_{n\geq 0} \frac{z^n}{(n+c)^s}\,, \tag{K.30}$$

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 (K.25) using (K.27) as

$$\zeta^{-1}(z,\beta) = \left(1 - ze^{\beta}\right)\left(1 - z(1 + \phi_L(ze^{\beta},\beta,1))\right) \tag{K.31}$$

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 (K.31). The root can come from any of the two factors. For large value of $\beta$ (low temperatures) the smallest root is determined from the $(1-ze^{\beta})$ factor, which gave the contribution of a single large drop. For small $\beta$ (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 $\beta_c$ through the equation

$$1 - e^{-\beta_c}(1 + \zeta_R(\beta_c)) = 0 \tag{K.32}$$

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 $\beta$ 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 $\beta$ 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 [19] 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 $\zeta^{-1}$ in (K.31), each of them is an analytic function and contains the smallest root within the radius of convergence of the series for the relevant $\beta$ values.

## K.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 of fractals of Hutchinson [23] and the iterated dynamical systems introduced by Barnsley and collaborators [12]. 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 K.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 case with multifractals. 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

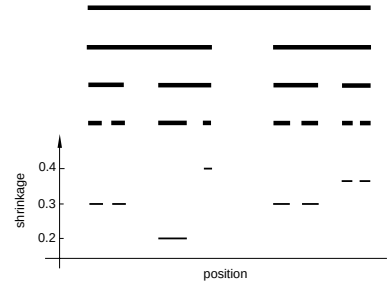



**Figure K.1:** Construction of the steps of the scaling function from a Cantor set. From one level to the next in the construction of the Cantor set the covers are shrunk, each parent segment into two children segments. The shrinkage of the last level of the construction is plotted and by removing the gaps one has an approximation to the scaling function of the Cantor set.

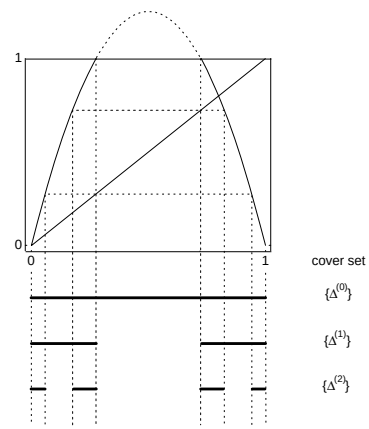



**Figure K.2:** A Cantor set presentation function. The Cantor set is the set of all points that under iteration do not leave the interval $[0, 1]$. This set can be found by backwards iterating the gap between the two branches of the map. The dotted lines can be used to find these backward images. At each step of the construction one is left with a set of segments that form a cover of the Cantor set.

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 K.2. Drawn is the map

$$F(x) = 5x(1 - x) \tag{K.33}$$

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 $F$, (K.33), it will hop between the branches and eventually get mapped to minus infinity. An orbit point is guaranteed to go to minus infinity if it lands in the gap. The hopping of the point defines the orbit of the initial point $x$: $x \mapsto x_1 \mapsto x_2 \mapsto \cdots$. 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 $t = 0, 1, 2, \ldots$.

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 [13].

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 K.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 K.2. As $n \to \infty$ 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 K.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

$$\Delta^{(2)}(0,1) = F_0^{-1}\left(F_1^{-1}\left(\Delta^{(0)}\right)\right) . \tag{K.34}$$

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:

$$\Delta^{(n)}(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) = F_{\epsilon_1}^{-1} \circ F_{\epsilon_2}^{-1} \circ \cdots F_{\epsilon_n}^{-1}\left(\Delta^{(0)}\right) .$$

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

$$\sigma^{(n)}(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) = \frac{|\Delta^{(n)}(\epsilon_1, \epsilon_2, \ldots, \epsilon_n)|}{|\Delta^{(n-1)}(\epsilon_2, \ldots, \epsilon_n)|} .$$

We can then arrange the ratios $\sigma^{(n)}(\epsilon_1, \epsilon_2, \ldots, \epsilon_n)$ next to each other as piecewise constant segments in increasing order of their binary label $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ so that the collection of steps scan the unit interval. As $n \to \infty$ this collection of steps will converge to the scaling function.

## K.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 $\phi$. 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 $\Omega_0$ that the spin $\sigma_0$ 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 $\iota$ that returns one for any configuration:

$$-\beta f(\beta) = \lim_{n\to\infty} \frac{1}{n} \ln \|\mathcal{L}^n \iota\|$$

and the thermodynamic average of an observable

$$\langle g \rangle = \lim_{n\to\infty} \frac{\|\mathcal{L}^n g\|}{\|\mathcal{L}^n \iota\|} .$$

Both relations hold for almost all configurations. These relations are part of 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 hows 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 $F_{\sigma_k}$ such that

$$F_{\sigma_k}(0) = 0$$

and

$$F_{\sigma_1}\circ F_{\sigma_2}\circ\cdots\circ F_{\sigma_n}(0) = \phi(+,\sigma_1,\sigma_2,\ldots,\sigma_n,-,-,\ldots)\,.$$

This is a formal relation that expresses how the interaction is to be converted into a dynamical systems. In most examples $F_{\sigma_k}$ 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 use 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 ) 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 $\Omega$. Then an observable $g$ is a function from the set of configurations $\Omega$ 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 $\eta_0$ to its beginning, $\eta \vee \sigma$, in which case $\eta$ is at site 0, $\omega_0$ 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, $\mathrm{tr}\,\mathcal{L}^n = Z_n$ 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 $\{x_1(\sigma), \ldots, x_1(\sigma)\}$. The idea is to choose the observables in such a way that from their values on a particular

configuration $\sigma$ the configuration can be reconstructed. We also introduce the interaction observables $h_{\sigma_0}$.

To geometrize spin systems, the interactions are assumed to be translationally invariant. The spins $\sigma_k$ will only assume a finite number of values. For simplicity, we will take the interaction $\phi$ among the spins to depend only on pairwise interactions,

$$\phi(\sigma) = \phi(\sigma_0, \sigma_1, \sigma_2, \ldots) = J_0\sigma_0 + \sum_{n>0} \delta_{\sigma_0,\sigma_n} J_1(n)\,, \tag{K.35}$$

and limit $\sigma_k$ 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 $J_1(n) = e^{-\alpha n}$. 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 [12] 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 $\sigma = \{\sigma_0, \sigma_1, \ldots\}$:

$$\mathcal{T}g(\sigma) = g(+ \vee \sigma)e^{-\beta\phi(+\vee\sigma)} + g(- \vee \sigma)e^{-\beta\phi(-\vee\sigma)}\,. \tag{K.36}$$

By $+ \vee \sigma$ we mean the configuration obtained by prepending $+$ to the beginning of $\sigma$ resulting in the configuration $\{+, \sigma_0, \sigma_1, \ldots\}$. When the range becomes infinite, $\mathrm{tr}\,\mathcal{T}^n$ 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 [13] 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

$$-\beta f(\beta) = \lim_{n\to\infty} \frac{1}{n} \ln \|\mathcal{T}^n u\|\,. \tag{K.37}$$

To construct a smooth dynamical system that reproduces the properties of $\mathcal{T}$, one uses the phase space reconstruction technique of Packard *et al.* [6] and Takens [7], and introduces a vector of state observables $x(\sigma) = \{x_1(\sigma), \ldots, x_D(\sigma)\}$.

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 $\sigma$, there exist an analytic function $G$ such that $G(x_1(\sigma), \ldots, x_D(\sigma)) = g(\sigma)$. 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 [14] 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

$$\mathcal{T} G\left(x(\sigma)\right) = \sum_{\eta \in \{+,-\}} G\left(x_+\left(\eta \vee \sigma\right), x_-\left(\eta \vee \sigma\right)\right) e^{-\beta x_\eta(\sigma)} . \tag{K.38}$$

In this equation the only reference to the configuration $\sigma$ 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 $F_\eta$ that take $x(\sigma)$ to $x(\eta \vee \sigma)$.

We can now define a new operator $\mathcal{L}$

$$\mathcal{L} G\left(x\right) \stackrel{\text{def}}{=} \mathcal{T} G(x(\sigma)) = \sum_{\eta \in \{+,-\}} G\left(F_\eta(x)\right) e^{-\beta x_\eta} , \tag{K.39}$$

where all dependencies on $\sigma$ 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 $F_\eta$. 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 $\phi$ 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 $\det\left(1 - z\mathcal{L}\right)$ 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 $\det(1 - z\mathcal{L}) = \exp(\mathrm{tr}\,\ln(1 - z\mathcal{L}))$. The logarithm is expanded into a power series and one is left with terms of the form $\mathrm{tr}\,\mathcal{L}^n$ to evaluate. For evaluating the trace, the $\mathcal{L}$ operator is equivalent to

$$\mathcal{L}G(x) = \int_{\mathbf{R}^D} dy\,\delta(y - f(x))e^{-\beta y}G(y) \tag{K.40}$$

from which the trace can be computed:

$$\mathrm{tr}\,\mathcal{L}^n = \sum_{x=f^{(\circ n)}(x)} \frac{e^{-\beta H(x)}}{|\det\,(1 - \partial_x f^{(\circ n)}(x))\,|} \tag{K.41}$$

with the sum running over all the fixed points of $f^{(\circ n)}$ (all spin configurations of a given length). Here $f^{(\circ n)}$ 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 $\mathrm{tr}\,\mathcal{L}^n$ we must compute the value of $\partial_x f^{(\circ n)}$; 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 $\epsilon$, such as $x = \phi(+, \epsilon_0, \epsilon_1, \ldots, -, -, \ldots)$. By choosing the sequence $\epsilon$ 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:

$$\partial_x f(x) = \lim_{m\to\infty} \frac{\phi(\epsilon \vee \eta^{(m)}) - \phi(\epsilon)}{\phi(\eta^{(m)})}\,, \tag{K.42}$$

where $\eta^{(m)}$ is a sequence of spin strings that make $\phi(\eta^{(m)})$ smaller and smaller. By substituting the definition of $\phi$ in terms of its pairwise interaction $J(n) = n^s a^{n^\gamma}$ and taking the limit for the sequences $\eta^{(m)} = \{+, -, -, \ldots, \eta_{m+1}, \eta_{m+2}, \ldots\}$ 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 $\gamma \to 0$.

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. Determining under which conditions the embedding is smooth is a complicated question [15]. 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

$$\phi(\sigma) = \sum_{n\geq 1} \delta_{\sigma_0,\sigma_n} \sum_k p_k(n) a_k^{n^\gamma} \tag{K.43}$$

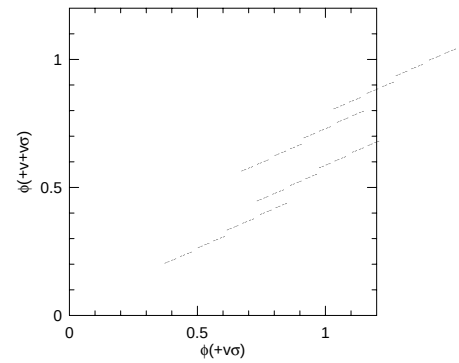



**Figure K.3:** The spin adding map $F_+$ for the potential $J(n) = \sum n^2 a^{\alpha n}$. The action of the map takes the value of the interaction energy between + and the semi-infinite configuration $\{\sigma_1, \sigma_2, \sigma_3, \ldots\}$ and returns the interaction energy between + and the configuration $\{+, \sigma_1, \sigma_2, \sigma_3, \ldots\}$.

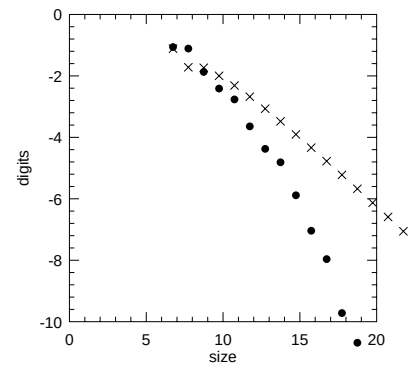



**Figure K.4:** Number of digits for the Fredholm method (•) and the transfer function method (×). The size refers to the largest cycle considered in the Fredholm expansions, and the truncation length in the case of the transfer matrix.

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 $x_{0,k}, x_{1,k}, \ldots$ 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 K.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 [17, 16]. We will illustrate this with the polynomial-exponential interactions (K.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 [17]. The convergence is illustrated in figure K.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 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 [17] 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.

## 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 $256 \times 256$ matrix. This make the method of cycle expansions practical for analytic calculations.

## Commentary

**Remark K.1** Presentation functions. The best place to read about Feigenbaum's work is in his review article published in *Los Alamos Science* (reproduced in various reprint collections and conference proceedings, such as ref. [5]). Feigenbaum's *Journal of Statistical Physics* article [13] is the easiest place to learn about presentation functions.

**Remark K.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 Sweneson-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:

$$\phi(\sigma_0, \sigma_1, \sigma_2, \ldots) = J_0\sigma_0 + \sum \delta(\sigma_0, \sigma_n) J_1(n) + \sum \delta(\sigma_0, \sigma_{n_1}, \sigma_{n_2}) J_2(n_1, n_2) + \cdots$$

where the $J_k$ are symmetric functions of their arguments and the $\delta$ are arbitrary discrete functions. This includes external constant fields ($J_0$), but it excludes site dependent fields such as a random external magnetic field.

## Exercises

K.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.

K.2. **Automaton for a droplet.** Find the Markov 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 (K.13) and (K.18).

Hint: the Markov graph is infinite.

K.3. **Spectral determinant for $a^n$ interactions** Compute the spectral determinant for 1-dimensional Ising model with the interaction

$$\phi(\sigma) = \sum_{k>0} a^k \delta(\sigma_0, \sigma_k) .$$

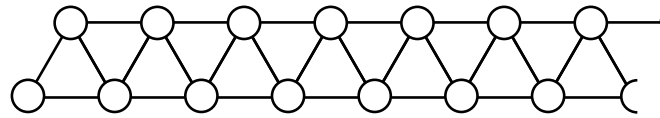
Take $a$ as a number smaller than $1/2$.

(a) What is the dynamical system this generates? That is, find $F_+$ and $F_-$ as used in (K.39).

(b) Show that

$$\frac{d}{dx} F_{\{+ \text{ or} -\}} = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$$

K.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.

K.5. **Infinite symbolic dynamics** Let $\sigma$ be a function that returns zero or one for every infinite binary string: $\sigma : \{0,1\}^{\mathbb{N}} \to \{0,1\}$. Its value is represented by $\sigma(\epsilon_1, \epsilon_2, \ldots)$ where the $\epsilon_i$ 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

$$\|a\| = \sup_{\{\epsilon_i\}} |a(\epsilon_1, \epsilon_2, \ldots)| < \infty .$$

For these functions

$$\mathcal{T} a(\epsilon_1, \epsilon_2, \ldots) = a(0, \epsilon_1, \epsilon_2, \ldots)\sigma(0, \epsilon_1, \epsilon_2, \ldots) + a(1, \epsilon_1,$$

The function $\sigma$ 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 $\mathcal{T}$:

$$\begin{aligned} T_n a(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) = \\ a(0, \epsilon_1, \epsilon_2, \ldots, \epsilon_{n-1})\sigma(0, \epsilon_1, \epsilon_2, \ldots, \epsilon_{n-1}) + \\ a(1, \epsilon_1, \epsilon_2, \ldots, \epsilon_{n-1})\sigma(1, \epsilon_1, \epsilon_2, \ldots, \epsilon_{n-1}) . \end{aligned}$$

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.

# References

[K.1] Ya. Sinai. Gibbs measures in ergodic theory. *Russ. Math. Surveys*, 166:21–69, 1972.

[K.2] R. Bowen. Periodic points and measure for axiom-A diffeomorphisms. *Transactions Amer. Math. Soc.*, 154:377–397, 1971.

[K.3] D. Ruelle. Statistical mechanics on a compound set with $Z^{\nu}$ action satisfying expansiveness and specification. *Transactions Amer. Math. Soc.*, 185:237–251, 1973.

[K.4] E. B. Vul, Ya. G. Sinai, and K. M. Khanin. Feigenbaum universality and the thermodynamic formalism. *Uspekhi Mat. Nauk.*, 39:3–37, 1984.

[K.5] M.J. Feigenbaum, M.H. Jensen, and I. Procaccia. Time ordering and the thermodynamics of strange sets: Theory and experimental tests. *Physical Review Letters*, 57:1503–1506, 1986.

[K.6] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw. Geometry from a time series. *Physical Review Letters*, 45:712 – 716, 1980.

[K.7] F. Takens, Detecting strange attractors in turbulence. In *Lecture Notes in Mathematics 898*, pages 366–381. Springer, Berlin, 1981.

[K.8] R. Mainieri. Thermodynamic zeta functions for Ising models with long range interactions. *Physical Review A*, 45:3580, 1992.

[K.9] R. Mainieri. Zeta function for the Lyapunov exponent of a product of random matrices. *Physical Review Letters*, 68:1965–1968, March 1992.

[K.10] D. Wintgen. Connection between long-range correlations in quantum spectra and classical periodic orbits. *Physical Review Letters*, 58(16):1589–1592, 1987.

[K.11] G. S. Ezra, K. Richter, G. Tanner, and D. Wintgen. Semiclassical cycle expansion for the Helium atom. *Journal of Physics B*, 24(17):L413–L420, 1991.

[K.12] H. A. Kramers and G. H. Wannier. Statistics of the two-dimensional ferromagnet. Part I. *Physical Review*, 60:252–262, 1941.

[K.13] D. Ruelle. Statistical mechanics of a one-dimensional lattice gas. *Communications of Mathematical Physics*, 9:267–278, 1968.

[K.14] David Ruelle. *Thermodynamic Formalism*. Addison-Wesley, Reading, 1978.

[K.15] P. Walters. *An introduction to ergodic theory*, volume 79 of *Graduate Text in Mathematics*. Springer-Verlag, New York, 1982.

[K.16] H.H. Rugh. *Time evolution and correlations in chaotic dynamical systems*. PhD thesis, Niels Bohr Institute, 1992.

[K.17] M.C. Gutzwiller. The quantization of a classically ergodic system. *Physica D*, 5:183–207, 1982.

[K.18] M. Feigenbaum. The universal metric properties of non-linear transformations. *Journal of Statistical Physics*, 19:669, 1979.

[K.19] G.A. Baker. One-dimensional order-disorder model which approaches a second order phase transition. *Phys. Rev.*, 122:1477–1484, 1961.

[K.20] M. E. Fisher. The theory of condensation and the critical point. *Physics*, 3:255–283, 1967.

[K.21] G. Gallavotti. Funzioni zeta ed insiemi basilari. *Accad. Lincei. Rend. Sc. fis. mat. e nat.*, 61:309–317, 1976.

[K.22] R. Artuso. Logarithmic strange sets. *J. Phys. A.*, 21:L923–L927, 1988.

[K.23] Hutchinson

# Appendix L

# Noise/quantum corrections

(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 classical 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 $D \to 0$. The effect of the noise can be taken into account by adding noise corrections to the classical 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.

## L.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 would like to 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 Fig. L.1).

The motion close to a periodic orbits is regular in both cases. This is due to the fact, that we can linearize the Hamiltonian close to an orbit, and linear systems

**Figure L.1:** Poincaré section close to a stable and an unstable periodic orbit

are always integrable. The linearized Hamilton's equations close to the periodic orbit $(q_p(t) + q, p_p(t) + p)$ look like

$$\dot{q} = +\partial^2_{pq} H(q_p(t), p_p(t))q + \partial^2_{pp} H(q_p(t), p_p(t))p, \tag{L.1}$$

$$\dot{p} = -\partial^2_{qq} H(q_p(t), p_p(t))q - \partial^2_{qp} H(q_p(t), p_p(t))p, \tag{L.2}$$

where the new coordinates $q$ and $p$ are relative to a periodic orbit. This linearized equation can be regarded as a $d$ dimensional 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

$$\partial_t H(q_p(t) + q, p_p(t) + p) = 0. \tag{L.3}$$

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

$$H_0(x_\|, p_\|, x_n, p_n) = \frac{1}{2}p_\|^2 + U(x_\|) + \sum_{n=1}^{d-1} \frac{1}{2}(p_n^2 \pm \omega_n^2 x_n^2), \tag{L.4}$$

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, $\omega$ is the Lyapunov exponent of the orbit

$$\omega_n = \ln \Lambda_{p,n}/T_p, \tag{L.5}$$

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 $\omega$ as

$$\Lambda_{p,n} = e^{-i\omega_n T_p}. \tag{L.6}$$

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

$$E_n = \hbar\omega_n\left(j_n + \frac{1}{2}\right) \text{for stable modes,} \tag{L.7}$$
$$E_n = -i\hbar\omega_n\left(j_n + \frac{1}{2}\right) \text{for unstable modes,}$$

where $j_n = 0, 1, \ldots$. 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 trough the classical action function of the mode:

$$\frac{1}{2\pi}\oint p_{\|}dx_{\|} = \frac{1}{2\pi}S_{\|}(E_m) = \hbar\left(m + \frac{m_p\pi}{2}\right), \tag{L.8}$$

where $m_p$ is the topological index of the motion in the parallel direction. This latter condition can be rewritten by a very useful trick into the equivalent form

$$(1 - e^{iS_{\|}(E_m)/\hbar - im_p\pi/2}) = 0. \tag{L.9}$$

The eigen-energies of a semiclassically quantized periodic orbit are all the possible energies

$$E = E_m + \sum_{n=1}^{d-1} E_n. \tag{L.10}$$

This relation allows us to change in (L.9) $E_m$ 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

$$\Delta_p(E) = \prod_{j_1,\ldots,j_{d-1}} (1 - e^{iS_{\|}(E - \sum_n \hbar s_n\omega_n(j_n+1/2))/\hbar - im_p\pi/2}). \tag{L.11}$$

If we Taylor expand the action around $E$ to first order

$$S_{\|}(E + \epsilon) \approx S_{\|}(E) + T(E)\epsilon, \tag{L.12}$$

where $T(E)$ is the period of the orbit, and use the relations of $\omega$ and the eigenvalues of the Jacobi matrix, we get the expression of the Selberg product

$$\Delta_p(E) = \prod_{j_1,\ldots,j_{d-1}} \left(1 - \frac{e^{iS_p(E)/\hbar - im_p\pi/2}}{\prod_n \Lambda_{p,n}^{(1/2+j_n)}}\right). \tag{L.13}$$

If we use the right convention for the square root we get exactly the $d$ dimensional expression of the Selberg product formula we derived from the Gutzwiller trace

formula in ? . 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 BS quantized energy of one of the periodic orbits, we have to take the product of these determinants:

$$\Delta(E) = \prod_p \Delta_p(E). \tag{L.14}$$

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 eigen=energies of the whole system !

So we learned, that both stable and unstable orbits are integrable systems and can be individually quantized semiclassically by the old Bohr-Sommerfeld rules. So we almost completed the program of Sommerfeld to quantize general systems with the method of Bohr. *Let us have a remark here. In addition to the Bohr-Sommerfeld rules, we used the unjustified approximation (L.12). Sommerfeld would never do this ! At that point we loose some important precision compared to the BS rules and we get somewhat worse results than a semiclassical formula is able to do. We will come back to this point later 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

$$H(x_\|, p_\|, x_n, p_n) = H_0(x_\|, p_\|, x_n, p_n) + H_A(x_\|, x_n, p_n), \tag{L.15}$$

where the anaharmonic part can be written as a sum of homogeneous polynomials of $x_n$ and $p_n$ with $x_\|$ dependent coefficients:

$$H_A(x_\|, x_n, p_n) = \sum_{k=3} H^k(x_\|, x_n, p_n) \tag{L.16}$$

$$H^k(x_\|, x_n, p_n) = \sum_{\sum l_n + m_n = k} H^k_{l_n, m_n}(x_\|) x_n^{l_n} p_n^{m_n} \tag{L.17}$$

This classical Hamiltonian is hopeless from Sommerfeld's point of view, since it is non integrable. However, Birkhoff in 1927[3] introduced the concept of normal form, which helps us out from this problem by giving successive integrable approximation to a non-integrable problem. Let's learn a bit more about it!

[3]It is really a pity, that in 1926 Schrödinger introduced the wave mechanics and blocked the development of Sommerfeld's concept.

## L.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 momentums $p_n$ the quadratic part of the Hamiltonian will be a set of oscillators

$$H_0(x_n, p_n) = \sum_{n=1}^{d} \frac{1}{2}(p_n^2 + \omega_n^2 x_n^2). \tag{L.18}$$

The full Hamiltonian can be rewritten with the new coordinates

$$H(x_n, p_n) = H_0(x_n, p_n) + H_A(x_n, p_n), \tag{L.19}$$

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

$$H_A(x_n, p_n) = \sum_{j=3}^{\infty} H^j(x_n, p_n), \tag{L.20}$$

$$H^j(x_n, p_n) = \sum_{|l|+|m|=j} h^j_{lm} x^l p^m, \tag{L.21}$$

where $h^j_{lm}$ are real constants and we used the multi-indices $l := (l_1, ..., l_d)$ with definitions

$$|l| = \sum l_n, x^l := x_1^{l_1} x_2^{l_2} ... x_d^{l_d}.$$

Birkhoff showed, that that by successive canonical transformations one can introduce new momentums 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

$$\tau_n = \frac{1}{2}(p_n^2 + \omega_n^2 x_n^2), \tag{L.22}$$

where $x_n$ and $p_n$ are the new coordinates and momentums, but $\omega_n$ is the original frequency. This is called the Birkhoff normal form of degree $N$:

$$H(x_n, p_n) = \sum_{j=2}^{N} H^j(\tau_1, ..., \tau_d), \tag{L.23}$$

where $H^j$ are homogeneous degree $j$ polynomials of $\tau$-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 e.g. 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 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 ($\sim 20$).

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 BS quantization.

## L.3 Bohr-Sommerfeld quantization of periodic orbits

There is some difference between equilibrium points and periodic orbits. The Hamiltonian (L.4) 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 (L.17) is

$$H(x_\|, p_\|, \tau_1, ...\tau_{d-1}) = H_0(x_\|, p_\|, \tau_1, ..., \tau_{d-1}) + \sum_{j=2}^{N} U^j(x_\|, \tau_1, ..., \tau_{d-1}), \quad \text{(L.24)}$$

where $U^j$ is a $j$th order homogeneous polynomial of $\tau$-s with $x_\|$ dependent coefficients. The orthogonal part can be BS quantized by quantizing the individual oscillators, replacing $\tau$-s as we did in (L.8). This leads to a one dimensional effective potential indexed by $j_1, ..., j_{d-1}$

$$H(x_\|, p_\|, j_1, ..., j_{d-1}) = \frac{1}{2}p_\|^2 + U(x_\|) + \sum_{n=1}^{d-1} \hbar s_n \omega_n (j_n + 1/2) + \quad \text{(L.25)}$$

$$+ \sum_{k=2}^{N} U^k(x_\|, \hbar s_1 \omega_1 (j_1 + 1/2), \hbar s_2 \omega_2 (j_2 + 1/2), ..., \hbar s_{d-1} \omega_{d-1} (j_{d-1} + 1/2)),$$

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 BS quantized for any given set of $j$-s

$$S_p(E, j_1, ..., j_{d-1}) = \oint p_{\|} dx_{\|} = \tag{L.26}$$
$$= \oint dx_{\|} \sqrt{E - \sum_{n=1}^{d-1} \hbar s_n \omega_n (j_n + 1/2) - U(x_{\|}, j_1, ..., j_{d-1})} = 2\pi\hbar(m + m_p/2),$$

where $U$ contains all the $x_{\|}$ dependent terms of the Hamiltonian. The spectral determinant becomes

$$\Delta_p(E) = \prod_{j_1,...,j_{d-1}} (1 - e^{iS_p(E,j_1,...,j_{d-1})/\hbar - m_p\pi/2}). \tag{L.27}$$

This expression completes the Sommerfeld method and tells us how to quantize chaotic or general Hamiltonian systems. Unfortunately, quantum mechanics postponed this nice formula until our book.

This formula has been derived with the help of the semiclassical Bohr-Sommerfeld quantization rule and the classical 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 attracting 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 sofar 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 classical quantities like $x^2p^2$. On the classical 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 (L.26) is not exact. We have to go back to the level of the Schrödinger equation if we would like to get the exact expression.

## L.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

$$\mathrm{Tr}G(x,x',t) = \int dxG(x,x,t) = \int \mathcal{D}xe^{iS(x,t)/\hbar}, \tag{L.28}$$

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 classical action calculated along the path. The trace in the saddle point calculation is a sum for classical periodic orbits and zero length orbits, since these are the extrema of the action $\delta S(x,t) = 0$ for closed paths:

$$\mathrm{Tr}G(x,x',t) = g_0(t) + \sum_{p\in PO} \int \mathcal{D}\xi_p e^{iS(\xi_p+x_p(t),t)/\hbar}, \tag{L.29}$$

where $g_0(t)$ is the zero length orbit contribution. We introduced the new coordinate $\xi_p$ 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. Sofar nothing new has been invented. If we continue the standard textbook calculation scheme, we have to Taylor expand the action in $\xi_p$ 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 don't compute the path sum directly. We use the correspondence between path integrals and partial differential equations. This idea comes from Maslov [5] and a good summary is in ref. [6]. We search for that Schrödinger equation, which leads to the path sum

$$\int \mathcal{D}\xi_p e^{iS(\xi_p+x_p(t),t)/\hbar}, \tag{L.30}$$

where the action around the periodic orbit is in a multi dimensional Taylor expanded form:

$$S(x,t) = \sum_{\mathbf{n}}^{\infty} s_{\mathbf{n}}(t)(x - x_p(t))^{\mathbf{n}}/\mathbf{n}!. \tag{L.31}$$

The symbol $\mathbf{n} = (n_1, n_2, ..., n_d)$ denotes the multi index in $d$ dimensions, $\mathbf{n}! = \prod_{i=1}^{d} n_i!$ 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

$$\partial_t S + \frac{1}{2}(\nabla S)^2 + U = 0, \tag{L.32}$$

in which the potential is expanded in a multidimensional Taylor series around the orbit

$$U(x) = \sum_{\mathbf{n}} u_{\mathbf{n}}(t)(x - x_p(t))^{\mathbf{n}}/\mathbf{n}!. \tag{L.33}$$

The Schrödinger equation

$$i\hbar\partial_t\psi = \hat{H}\psi = -\frac{\hbar^2}{2}\Delta\psi + U\psi, \tag{L.34}$$

with this potential also can be expanded around the periodic orbit. Using the WKB ansatz

$$\psi = \varphi e^{iS/\hbar}, \tag{L.35}$$

we can construct a Schrödinger equation corresponding to a given order of the Taylor expansion of the classical action. The Schrödinger equation induces the Hamilton-Jacobi equation (L.32) for the phase and the transport equation of Maslov and Fjedoriuk [7] for the amplitude:

$$\partial_t\varphi + \nabla\varphi\nabla S + \frac{1}{2}\varphi\Delta S - \frac{i\hbar}{2}\Delta\varphi = 0. \tag{L.36}$$

This is the partial differential equation, solved in the neighborhood of a periodic orbit with the expanded action (L.31), which belongs to the local path-sum (L.30).

If we know the Green's function $G_p(\xi, \xi', t)$ corresponding to the local equation (L.36), then the local path sum can be converted back into a trace:

$$\int \mathcal{D}\xi_p e^{i/\hbar \sum_{\mathbf{n}} S_{\mathbf{n}}(x_p(t),t)\xi_p^{\mathbf{n}}/\mathbf{n}!} = \mathrm{Tr}G_p(\xi, \xi', t). \tag{L.37}$$

The saddle point expansion of the trace in terms of local traces then becomes

$$\mathrm{Tr}G(x, x', t) = \mathrm{Tr}G_W(x, x', t) + \sum_p \mathrm{Tr}G_p(\xi, \xi', t), \tag{L.38}$$

where $G_W(x, x', t)$ denotes formally the Green's function expanded around zero length (non moving) periodic orbits, known as the Weyl term [8]. Each Green's function can be Fourier-Laplace transformed independently and by definition we get in the energy domain:

$$\mathrm{Tr}G(x, x', E) = g_0(E) + \sum_p \mathrm{Tr}G_p(\xi, \xi', E). \tag{L.39}$$

Notice, 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. indexGreen's function!energy dependent

The spectral determinant is a function which has zeroes at the eigen-energies $E_n$ 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:

$$\mathrm{Tr}G(x, x', E) = \sum_n \frac{1}{E - E_n} = \frac{d}{dE} \log \Delta(E). \tag{L.40}$$

We can define the spectral determinant $\Delta_p(E)$ also for the local operators and we can write

$$\mathrm{Tr}G_p(\xi, \xi', E) = \frac{d}{dE} \log \Delta_p(E). \tag{L.41}$$

Using (L.39) 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 classical energy of the periodic orbit coincides with an eigen-energy $E$ of the local Schrödinger equation around the periodic orbit, then the corresponding stationary eigenfunction fulfills

$$\psi_p(\xi, t + T_p) = \int d\xi' G_p(\xi, \xi', t + T_p)\psi_p(\xi', t) = e^{-iET_p/\hbar}\psi_p(\xi, t), \tag{L.42}$$

where $T_p$ is the period of the prime orbit $p$. If the classical energy of the periodic orbit is not an eigen=energy of the local Schrödinger equation, the non-stationary eigenfunctions fulfill

$$\psi_p^{\mathbf{l}}(\xi, t+T_p) = \int d\xi' G_p(\xi, \xi', t+T_p)\psi_p(\xi', t) = e^{-iET_p/\hbar}\lambda_p^{\mathbf{l}}(E)\psi_p^{\mathbf{l}}(t), \quad \text{(L.43)}$$

where $\mathbf{l} = (l_1, l_2, ...)$ is a multi-index of the possible quantum numbers of the local Schrödinger equation. If the eigenvalues $\lambda_p^{\mathbf{l}}(E)$ are known the local functional determinant can be written as

$$\Delta_p(E) = \prod_{\mathbf{l}}(1 - \lambda_p^{\mathbf{l}}(E)), \quad \text{(L.44)}$$

since $\Delta_p(E)$ is zero at the eigen=energies of the local Schrödinger problem. We can insert the ansatz (L.35) and reformulate (L.43) as

$$e^{\frac{i}{\hbar}S(t+T_p)}\varphi_p^{\mathbf{l}}(t+T_p) = e^{-iET_p/\hbar}\lambda_p^{\mathbf{l}}(E)e^{\frac{i}{\hbar}S(t)}\varphi_p^{\mathbf{l}}(t). \quad \text{(L.45)}$$

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

$$S_p(E) = \oint pdq = \int_0^{T_p} L(t)dt + ET_p, \quad \text{(L.46)}$$

we get

$$e^{\frac{i}{\hbar}S_p(E)}\varphi_p^{\mathbf{l}}(t+T_p) = \lambda_p^{\mathbf{l}}(E)\varphi_p^{\mathbf{l}}(t). \quad \text{(L.47)}$$

Introducing the eigen-equation for the amplitude

$$\varphi_p^{\mathbf{l}}(t+T_p) = R_{\mathbf{l},p}(E)\varphi_p^{\mathbf{l}}(t), \quad \text{(L.48)}$$

the local spectral determinant can be expressed as a product for the quantum numbers of the local problem:

$$\Delta_p(E) = \prod_{\mathbf{l}}(1 - R_{\mathbf{l},p}(E)e^{\frac{i}{\hbar}S_p(E)}). \quad \text{(L.49)}$$

Since $\hbar$ is a small parameter we can develop a perturbation series for the amplitudes $\varphi_p^{\mathbf{l}}(t) = \sum_{m=0}^{\infty}\left(\frac{i\hbar}{2}\right)^m \varphi_p^{\mathbf{l}(m)}(t)$ which can be inserted into the equation (L.36) and we get an iterative scheme starting with the semiclassical solution $\varphi^{\mathbf{l}(0)}$:

$$\begin{aligned} \partial_t\varphi^{\mathbf{l}(0)} + \nabla\varphi^{\mathbf{l}(0)}\nabla S + \frac{1}{2}\varphi^{\mathbf{l}(0)}\Delta S &= 0, \quad \text{(L.50)} \\ \partial_t\varphi^{\mathbf{l}(m+1)} + \nabla\varphi^{\mathbf{l}(m+1)}\nabla S + \frac{1}{2}\varphi^{\mathbf{l}(m+1)}\Delta S &= \Delta\varphi^{\mathbf{l}(m)}. \end{aligned}$$

The eigenvalue can also be expanded in powers of $i\hbar/2$:

$$R_{\mathbf{l},p}(E) = \exp\left\{\sum_{m=0}^{\infty}\left(\frac{i\hbar}{2}\right)^m C_{\mathbf{l},p}^{(m)}\right\} \tag{L.51}$$

$$= \exp(C_{\mathbf{l},p}^{(0)})\{1 + \frac{i\hbar}{2}C_{\mathbf{l},p}^{(1)} + \left(\frac{i\hbar}{2}\right)^2\left(\frac{1}{2}(C_{\mathbf{l},p}^{(1)})^2 + C_{\mathbf{l},p}^{(2)}\right) + \dots . \tag{L.52}$$

The eigenvalue equation (L.48) in $\hbar$ expanded form reads as

$$\begin{aligned}
\varphi_p^{\mathbf{l}(0)}(t+T_p) &= \exp(C_{\mathbf{l},p}^{(0)})\varphi_p^{\mathbf{l}(0)}(t), \\
\varphi_p^{\mathbf{l}(1)}(t+T_p) &= \exp(C_{\mathbf{l},p}^{(0)})[\varphi_p^{\mathbf{l}(1)}(t) + C_{\mathbf{l},p}^{(1)}\varphi_p^{\mathbf{l}(0)}(t)], \\
\varphi_p^{\mathbf{l}(2)}(t+T_p) &= \exp(C_{\mathbf{l},p}^{(0)})[\varphi_p^{\mathbf{l}(2)}(t) + C_{\mathbf{l},p}^{(1)}\varphi_p^{\mathbf{l}(1)}(t) + (C_{\mathbf{l},p}^{(2)} + \frac{1}{2}(C_{\mathbf{l},p}^{(1)})^2)\varphi_p^{\mathbf{l}(0)}(t)]
\end{aligned} \tag{L.53}$$

and so on. These equations are the conditions selecting the eigenvectors and eigenvalues and they hold for all $t$.

It is very convenient to expand the functions $\varphi_p^{\mathbf{l}(m)}(x,t)$ in Taylor series around the periodic orbit and to solve the equations (L.51) in this basis [10], since only a couple of coefficients should be computed to derive the first corrections. This technical part we are going to publish elsewhere [9]. One can derive in general the zero order term $C_{\mathbf{l}}^{(0)} = i\pi\nu_p + \sum_{i=1}^{d-1}\left(l_i + \frac{1}{2}\right)u_{p,i}$, where $u_{p,i} = \log\Lambda_{p,i}$ are the logarithms of the eigenvalues of the monodromy matrix $M_p$ and $\nu_p$ is the topological index of the periodic orbit. The first correction is given by the integral

$$C_{\mathbf{l},p}^{(1)} = \int_0^{T_p} dt \frac{\Delta\varphi_p^{\mathbf{l}(0)}(t)}{\varphi_p^{\mathbf{l}(0)}(t)}.$$

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 (L.36) behaves discontinuously when the trajectory $x_p(t)$ hits the wall. For the simplicity we consider a two dimensional billiard system here. The wave function on the wall before the bounce ($t_{-0}$ ) is given by

$$\psi_{in}(x,y(x),t) = \varphi(x,y(x),t_{-0})e^{iS(x,y(x),t_{-0})/\hbar}, \tag{L.54}$$

where $y(x) = Y_2x^2/2! + Y_3x^3/3! + Y_4x^4/4! + \dots$ 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

$$\psi_{out}(x,y(x),t) = \varphi(x,y(x),t_{+0})e^{iS(x,y(x),t_{+0})/\hbar}. \tag{L.55}$$

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

$$S(x, y(x), t_{-0}) = S(x, y(x), t_{+0}), \tag{L.56}$$

and

$$\varphi(x, y(x), t_{-0}) = -\varphi(x, y(x), t_{+0}). \tag{L.57}$$

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:

$$\Delta(E) = e^{W(E)} \prod_p \prod_{\mathbf{l}} (1 - R_{\mathbf{l},p}(E) e^{\frac{i}{\hbar} S_p(E)}). \tag{L.58}$$

This expression is the quantum generalization of the semiclassical Selberg-product formula [11]. A similar decomposition has been found for quantum Baker maps in ref. [12]. The functions

$$\zeta_{\mathbf{l}}^{-1}(E) = \prod_p (1 - R_{\mathbf{l},p}(E) e^{\frac{i}{\hbar} S_p(E)}) \tag{L.59}$$

are the generalizations of the Ruelle type [23] zeta functions. The trace formula can be recovered from (L.40):

$$\mathrm{Tr}G(E) = g_0(E) + \frac{1}{i\hbar} \sum_{p,\mathbf{l}} (T_p(E) - i\hbar \frac{d \log R_{\mathbf{l},p}(E)}{dE}) \frac{R_{\mathbf{l},p}(E) e^{\frac{i}{\hbar} S_p(E)}}{1 - R_{\mathbf{l},p}(E) e^{\frac{i}{\hbar} S_p(E)}}. \tag{L.60}$$

We can rewrite the denominator as a sum of a geometric series and we get

$$\mathrm{Tr}G(E) = g_0(E) + \frac{1}{i\hbar} \sum_{p,r,\mathbf{l}} (T_p(E) - i\hbar \frac{d \log R_{\mathbf{l},p}(E)}{dE}) (R_{\mathbf{l},p}(E))^r e^{\frac{i}{\hbar} r S_p(E)}. \tag{L.61}$$

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.

**Figure L.2:** A typical bounce on a billiard wall. The wall can be characterized by the local expansion $y(x) = Y_2x^2/2! + Y_3x^3/3! + Y_4x^4/4! + ....$

The semiclassical trace formula can be recovered by dropping the sub-leading term $-i\hbar d \log R_{\mathbf{l},p}(E)/dE$ and using the semiclassical eigenvalue $R^{(0)}_{\mathbf{l},p}(E) = e^{C^{\mathbf{l}(0)}_p} = e^{-i\nu_p\pi}e^{-\sum_i(l_i+1/2)u_{p,i}}$. Summation for the indexes $l_i$ yields the celebrated semiclassical amplitude

$$\sum_{\mathbf{l}}(R^{(0)}_{\mathbf{l},p}(E))^r = \frac{e^{-ir\nu_p\pi}}{|\det(\mathbf{1} - M^r_p)|^{1/2}}. \tag{L.62}$$

To have an impression about the improvement caused by the quantum corrections we have developed a numerical code [13] which calculates the first correction $C^{(1)}_{p,l}$ for general two dimensional 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 $C^{(1)}_{p,0}$ for the periodic orbits of the 3-disk scattering system [14] where the quantum corrections have been We have found agreement up to the fifth decimal digit, while our method generates these numbers with any desired precision. Unfortunately, the $l \neq 0$ coefficients cannot be compared to ref. [15], since the $l$ dependence was not realized there due to the lack of general formulas (L.58) and (L.59). However, the $l$ dependence can be checked on the 2 disk scattering system [16]. On the standard example [14, 15, 16, 18], when the distance of the centers ($R$) is 6 times the disk radius ($a$), we got

$$C^{(1)}_l = \frac{1}{\sqrt{2E}}(-0.625l^3 - 0.3125l^2 + 1.4375l + 0.625).$$

For $l = 0$ and 1 this has been confirmed by A. Wirzba [17], who was able to compute $C^{(1)}_0$ from his exact quantum calculation. Our method makes it possible to utilize the symmetry reduction of Cvitanović and Eckhardt and to repeat the fundamental domain cycle expansion calculation of ref. [18] 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 I. shows the numerical values of the exact quantum calculation [16], the semiclassical cycle expansion [10] and our corrected calculation. One can see, that the error of the corrected calculation vs. the error of the semiclassical calculation decreases with the wave-number. Besides the improved results, a fast convergence up to six decimal digits can be observed, which is just three decimal digits in the full domain calculation [15].

## References

[L.1] M. C. Gutzwiller, J. Math. Phys. **12**, 343 (1971); *Chaos in Classical and Quantum Mechanics* (Springer-Verlag, New York, 1990)

**Table L.1:** Real part of the resonances (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 $\delta_{SC}$ divided by the error of the first correction $\delta_{Corr}$. 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 |

[L.2] A. Selberg, J. Indian Math. Soc. **20**, 47 (1956)

[L.3] See examples in : CHAOS **2** (1) Thematic Issue; E. Bogomolny and C. Schmit, Nonlinearity **6**, 523 (1993)

[L.4] R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948)

[L.5] We thank E. Bogomolny for bringing this reference to our attention.

[L.6] V. M. Babić and V. S. Buldyrev, *Short Wavelength Diffraction Theory*, Springer Series on Wave Phenomena, Springer-Verlag (1990)

[L.7] V. P. Maslov and M. V. Fjedoriuk, *Semiclassical Approximation in Quantum Mechanics*, Dordrecht-Reidel (1981)

[L.8] R. B. Balian and C. Bloch, Ann. Phys. (New York) **60**, 81 (1970);ibid. **63**, 592 (1971); M.V. Berry, M.V., C.J. Howls, C.J. Proceedings of the Royal Society of London. **447**, 1931 (1994)

[L.9] P. E. Rosenqvist and G. Vattay, in progress.

[L.10] P. Cvitanović, P. E. Rosenqvist, G. Vattay and H. H. Rugh, CHAOS **3** (4), 619 (1993)

[L.11] A. Voros, J. Phys. **A21**, 685 (1988)

[L.12] A. Voros, Prog. Theor. Phys. Suppl. **116**,17 (1994); M. Saraceno and A. Voros, to appear in Physica D.

[L.13] The FORTRAN code is available upon e-mail request to G. Vattay.

[L.14] P. Gaspard and S. A. Rice, J. Chem. Phys. **90** 2225, 2242, 2255 (1989) **91** E3279 (1989)

[L.15] D. Alonso and P. Gaspard, Chaos **3**, 601 (1993); P. Gaspard and D. Alonso, Phys. Rev. **A47**, R3468 (1993)

[L.16] A. Wirzba, CHAOS **2**, 77 (1992); Nucl. Phys. **A560**, 136 (1993)

[L.17] A. Wirzba, private communication.

[L.18] P. Cvitanović and B. Eckhardt, Phys. Rev. Lett. **63**, 823 (1989)

# Appendix T

# 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 Markov 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 length 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 intersting quantity, such as the conductance
7. compute some number of the classical and/or quantum eigenvalues, if appropriate

## T.1 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-$d$ dynamical systems, chains of piecewise linear maps, where all transport coefficients can be evaluated analytically. The translational symmetry (24.10) 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

$$\hat{f}(x) = \begin{cases} \Lambda x & x \in [0, 1/4 + 1/4\Lambda] \\ -\Lambda x + (\Lambda + 1)/2 & x \in [1/4 + 1/4\Lambda, 3/4 - 1/4\Lambda] \\ \Lambda x + (1 - \Lambda) & x \in [3/4 - 1/4\Lambda, 1] \end{cases} . \tag{T.1}$$

The corresponding circle map $f(x)$ is obtained by modulo the integer part. The elementary cell map $f(x)$ is sketched in figure T.1. The map has the symmetry property

$$\hat{f}(\hat{x}) = -\hat{f}(-\hat{x}) , \tag{T.2}$$

so that the dynamics has no drift, and all odd derivatives of the generating function (24.3) with respect to $\beta$ evaluated at $\beta = 0$ vanish.

The cycle weights are given by

$$t_p = z^{n_p} \frac{e^{\beta \hat{n}_p}}{|\Lambda_p|} . \tag{T.3}$$

The diffusion constant formula for 1-$d$ maps is

$$D = \frac{1}{2} \frac{\left\langle \hat{n}^2 \right\rangle_\zeta}{\langle n \rangle_\zeta} \tag{T.4}$$

where the "mean cycle time" is given by

$$\langle n \rangle_\zeta = z \frac{\partial}{\partial z} \frac{1}{\zeta(0, z)} \bigg|_{z=1} = -\sum{}' (-1)^k \frac{n_{p_1} + \cdots + n_{p_k}}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|} , \tag{T.5}$$

the mean cycle displacement squared by

$$\left\langle \hat{n}^2 \right\rangle_\zeta = \frac{\partial^2}{\partial \beta^2} \frac{1}{\zeta(\beta, 1)} \bigg|_{\beta=0} = -\sum{}' (-1)^k \frac{(\hat{n}_{p_1} + \cdots + \hat{n}_{p_k})^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|} , \tag{T.6}$$

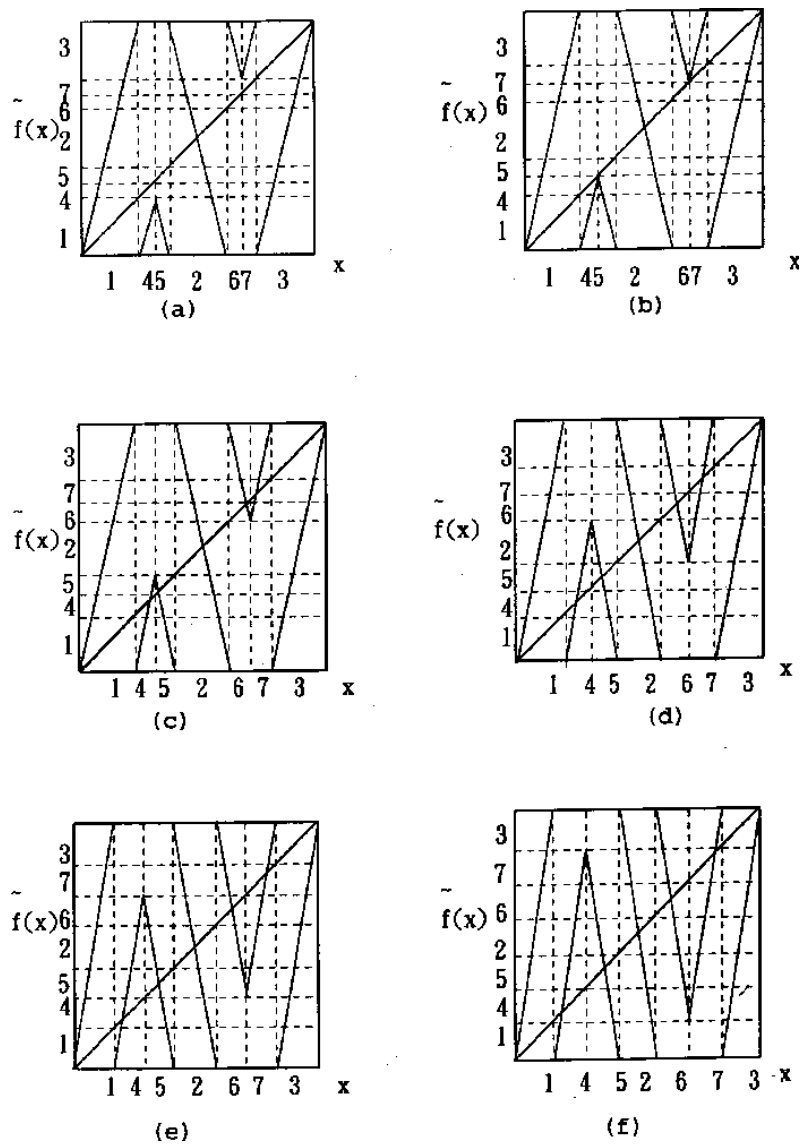



**Figure T.1:** (a)-(f) The sawtooth map (T.1) for the 6 values of parameter $a$ for which the folding point of the map aligns with the endpoint of one of the 7 intervals and yields a finite Markov partition (from ref. [1]). The corresponding Markov graphs are given in figure T.2.

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 (24.35) is convenient, but not essential for this project, so if you find sect. 19.1.1 too long a read, skip the symmetrization.

### T.1.1 The full shift

Take the map (T.1) and extend it to the real line. As in example of figure 24.3, denote by $a$ the critical value of the map (the maximum height in the unit cell)

$$a = \hat{f}(\frac{1}{4} + \frac{1}{4\Lambda}) = \frac{\Lambda + 1}{4} \,. \tag{T.7}$$

Describe the symbolic dynamics that you obtain when $a$ is an integer, and derive the formula for the diffusion constant:

$$D = \frac{(\Lambda^2 - 1)(\Lambda - 3)}{96\Lambda} \qquad \text{for } \Lambda = 4a - 1, \ a \in \mathbb{Z}. \tag{T.8}$$

If you are going strong, derive also the fromula for the half-integer $a = (2k+1)/2$, $\Lambda = 4a + 1$ case and email it to DasBuch@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 (24.21). [exercise 24.1]

### T.1.2 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 $1 \le a \le 2$. 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 T.1 (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 (10.2) is given by

$$\phi' = T\phi = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_4 \\ \phi_5 \\ \phi_2 \\ \phi_6 \\ \phi_7 \\ \phi_3 \end{pmatrix} \tag{T.9}$$

and the dynamics is unrestricted in the alphabet

$$\{1, \underline{41}, \underline{51}, 2, \underline{63}, \underline{73}, 3, \}.$$

One could diagonalize (T.9) on the computer, but, as we saw in sect. 10.4, the Markov graph figure T.2 (b) corresponding to figure T.1 (a) offers more insight into the dynamics. The dynamical zeta function

$$\begin{aligned} 1/\zeta &= 1 - (t_1 + t_2 + t_3) - 2(t_{14} + t_{37}) \\ 1/\zeta &= 1 - 3\frac{z}{\Lambda} - 4\cosh\beta\frac{z^2}{\Lambda^2}. \end{aligned} \tag{T.10}$$

follows from the loop expansion (13.13) of sect. 13.3.

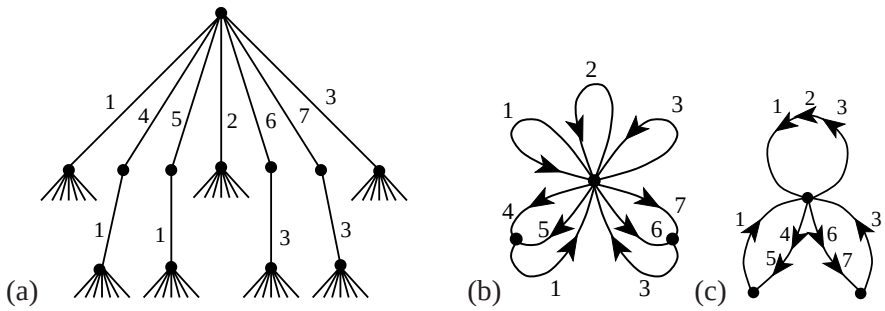



**Figure T.2:** (a) The sawtooth map (T.1) partition tree for figure T.1 (a); while intervals $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ map onto the whole unit interval, $f(\mathcal{M}_1) = f(\mathcal{M}_2) = f(\mathcal{M}_3) = \mathcal{M}$, intervals $\mathcal{M}_4, \mathcal{M}_5$ map onto $\mathcal{M}_1$ 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 three are similar to the whole tree and can be identified (as, for example, in figure 10.13), yielding (b) the Markov graph for the Markov partition of figure T.1 (a). (c) the Markov graph in the compact notation of (24.26).

The material flow conservation sect. 20.3 and the symmetry factorization (24.35) yield

$$0 = \frac{1}{\zeta(0,1)} = \left(1 + \frac{1}{\Lambda}\right)\left(1 - \frac{4}{\Lambda}\right)$$

which indeed is satisfied by the given value of $\Lambda$. 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 $\Lambda$. For more complicated transition matrices the factorization (24.35) is very helpful in reducing the order of the polynomial condition that fixes $\Lambda$.

The diffusion constant follows from (24.36) and (T.4)

$$\langle n \rangle_\zeta = -\left(1 + \frac{1}{\Lambda}\right)\left(-\frac{4}{\Lambda}\right), \quad \left\langle \hat{n}^2 \right\rangle_\zeta = \frac{4}{\Lambda^2}$$

$$D = \frac{1}{2}\frac{1}{\Lambda + 1} = \frac{1}{10}$$

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 T.1 and determine for what value of the parameter $a$ each of them is realized. Work out the Markov graph, symmetrization factorization and the diffusion constant, and check the material flow conservation for each case. Derive the diffusion constants listed in table T.1. It is not clear why the final answers tend to be so simple. Numerically, the case of figure T.1 (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 T.1, figure T.1 and (T.8)) 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

| figure T.1 | $\Lambda$ | $D$ |
|---|---|---|
| | 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 T.1:** The diffusion constant as function of the slope $\Lambda$ for the $a = 1, 2$ values of (T.8) and the 6 Markov partitions of figure T.1

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 10.13). 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.

### T.1.3 Diffusion coefficient, numerically

*(optional:)*
Attempt a numerical evaluation of

$$D = \frac{1}{2}\lim_{n\to\infty}\frac{1}{n}\left\langle \hat{x}_n^2 \right\rangle . \tag{T.11}$$

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 $\Lambda$; perhaps blow up a small region to show that the dependance of $D$ on the parameter $\Lambda$ is fractal. Compare with figure 24.5 and figures in refs. [1, 2, 8, 9].

### T.1.4 $D$ is a nonuniform function of the parameters

*(optional:)*
The dependence of $D$ on the map parameter $\Lambda$ is rather unexpected - even though

for larger $\Lambda$ 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 (T.1) for a random "generic" value of the parameter $\Lambda$, for example $\Lambda = 6$. The idea is to bracket this value of $\Lambda$ 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. 13.6 and appendix D.1), this is too ambitious for a week-long exam.

## References

[T.1] H.-C. Tseng, H.-J. Chen, P.-C. Li, W.-Y. Lai, C.-H. Chou and H.-W. Chen, "Some exact results for the diffusion coefficients of maps with pruned cycles," *Phys. Lett.* **A** 195, 74 (1994).

[T.2] C.-C. Chen, "Diffusion Coefficient of Piecewise Linear Maps," *Phys. Rev.* **E51**, 2815 (1995).

[T.3] H.-C. Tseng and H.-J. Chen, "Analytic results for the diffusion coefficient of a piecewise linear map," *Int. J. Mod. Phys.***B 10**, 1913 (1996).

## T.2 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-$d$ dynamical systems, chains of piecewise linear maps (24.9). The translational symmetry (24.10) 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 24.3. The map has the symmetry property

$$\hat{f}(\hat{x}) = -\hat{f}(-\hat{x}), \tag{T.12}$$

so that the dynamics has no drift, and all odd derivatives of the generating function (24.3) with respect to $\beta$ evaluated at $\beta = 0$ vanish.

The cycle weights are given by

$$t_p = z^{n_p} \frac{e^{\beta \hat{n}_p}}{|\Lambda_p|} . \tag{T.13}$$

The diffusion constant formula for 1-$d$ maps is

$$D = \frac{1}{2} \frac{\left\langle \hat{n}^2 \right\rangle_\zeta}{\langle n \rangle_\zeta} \tag{T.14}$$

where the "mean cycle time" is given by

$$\langle n \rangle_\zeta = z \frac{\partial}{\partial z} \frac{1}{\zeta(0,z)} \bigg|_{z=1} = -\sum{}' (-1)^k \frac{n_{p_1} + \cdots + n_{p_k}}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|} , \tag{T.15}$$

the mean cycle displacement squared by

$$\left\langle \hat{n}^2 \right\rangle_\zeta = \frac{\partial^2}{\partial \beta^2} \frac{1}{\zeta(\beta,1)} \bigg|_{\beta=0} = -\sum{}' (-1)^k \frac{(\hat{n}_{p_1} + \cdots + \hat{n}_{p_k})^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|} , \tag{T.16}$$

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.

### T.2.1 The full shift

Reproduce the formulas of sect. 24.2.1 for the diffusion constant $D$ for $\Lambda$ both even and odd integer.

| figure 24.4 | $\Lambda$ | $D$ |
|---|---|---|
| | 4 | $\frac{1}{4}$ |
| (a) | $2+\sqrt{6}$ | $1-\frac{3}{4}\sqrt{6}$ |
| (b) | $2\sqrt{2}+2$ | $\frac{15+2\sqrt{2}}{16+4\sqrt{2}}$ |
| (c) | 5 | 1 |
| (d) | $3+\sqrt{5}$ | $\frac{5}{2}\frac{\Lambda-1}{3\Lambda-4}$ |
| (e) | $3+\sqrt{7}$ | $\frac{5\Lambda-4}{3\Lambda-2}$ |
| | 6 | $\frac{5}{6}$ |

**Table T.2:** The diffusion constant as function of the slope $\Lambda$ for the $\Lambda = 4, 6$ values of (24.20) and the 5 Markov partitions like the one indicated in figure 24.4.

### T.2.2 Subshifts of finite type

We now work out examples 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.

Start by reproducing the formula (24.28) of sect. 24.2.3 for the diffusion constant $D$ for the Markov partition, the case where the critical point is mapped onto the right border of $I_{1_+}$.

Think up other non-integer values of the parameter $\Lambda$ for which the symbolic dynamics is given in terms of Markov partitions: in particular consider the remaining four cases for which the critical point is mapped onto a border of a partition in one iteration. Work out the Markov graph symmetrization factorization and the diffusion constant, and check the material flow conservation for each case. Fill in the diffusion constants missing in table T.2. It is not clear why the final answers tend to be so simple. What value of $\Lambda$ appears to yield the maximal diffusion constant?

The 7 cases considered here (see table T.2 and figure 24.4) are the 7 simplest complete Markov partitions in the $4 \leq \Lambda \leq 6$ interval, the criterion being that the critical points map onto partition boundary points. In case at hand, the "critical" point is the highest point of the left branch of the map (symmetry automatically takes care of the other critical point, the lowest point of the left branch), 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 6 primary intervals can be subdivided into 6 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.

### T.2.3 Diffusion coefficient, numerically

*(optional:)*

Attempt a numerical evaluation of

$$D = \frac{1}{2} \lim_{n\to\infty} \frac{1}{n} \left\langle \hat{x}_n^2 \right\rangle . \tag{T.17}$$

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 $\Lambda$; perhaps blow up a small region to show that the dependance of $D$ on the parameter $\Lambda$ is fractal. Compare with figure 24.5 and figures in refs. [1, 2, 8, 9].

### T.2.4 $D$ is a nonuniform function of the parameters

*(optional:)*
The dependence of $D$ on the map parameter $\Lambda$ is rather unexpected - even though for larger $\Lambda$ more points are mapped outside the unit cell in one iteration, the diffusion constant does not necessarily grow. Figure 24.5 taken from ref. [8] illustrates the fractal dependence of diffusion constant on the map parameter. An interpretation of this lack of monotonicity would be interesting.

You can also try applying periodic orbit theory to the sawtooth map (24.9) for a random "generic" value of the parameter $\Lambda$, for example $\Lambda = 4.5$. The idea is to bracket this value of $\Lambda$ by the nearby ones, for which higher and higher iterates of the critical value $a = \Lambda/2$ 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 $\Lambda = 4.5$. Judging how difficult such problem is already for a tent map (see sect. 13.6 and appendix D.1), this is too ambitious for a week-long exam.