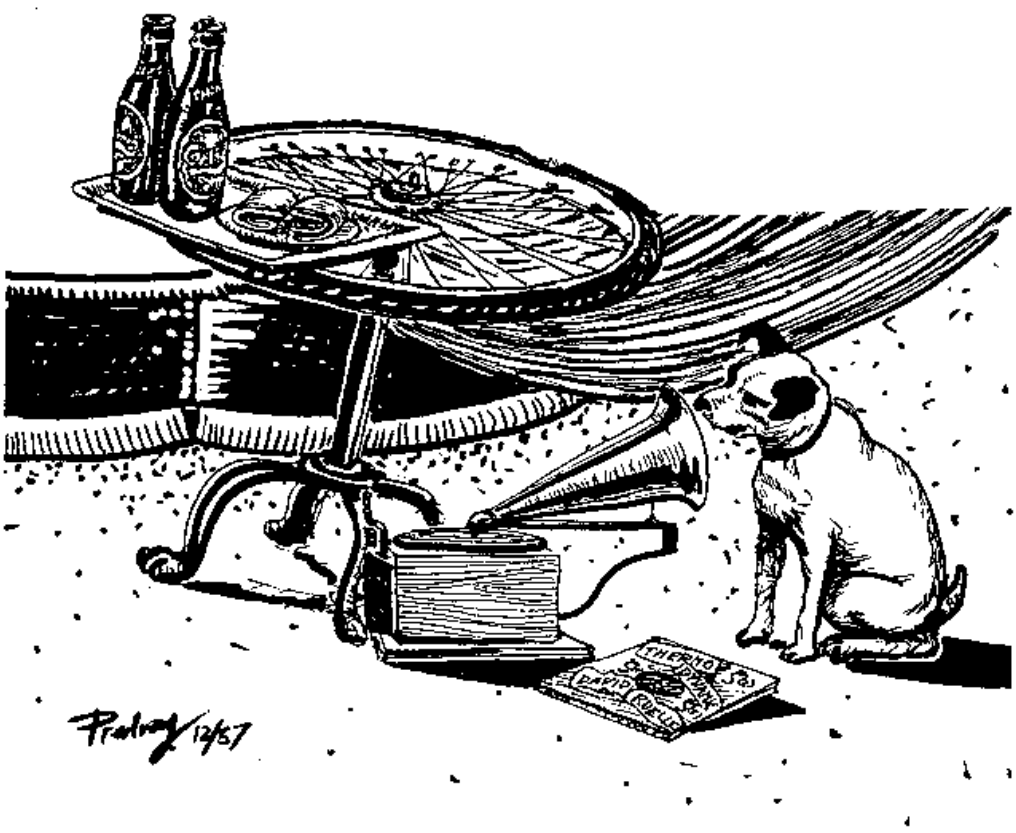
Classical and Quantum Chaos



Predrag Cvitanović – Roberto Artuso – Per Dahlqvist – Ronnie Mainieri – Gregor Tanner – Gábor Vattay – Niall Whelan – Andreas Wirzba

version 9.2.3 **Feb 26 2002**

printed June 19, 2002

`www.nbi.dk/ChaosBook/` comments to: predrag@nbi.dk

Contents

Contributors . . . x

1 Overture **1**
1.1 Why this book? . . . 2
1.2 Chaos ahead . . . 3
1.3 A game of pinball . . . 4
1.4 Periodic orbit theory . . . 13
1.5 Evolution operators . . . 18
1.6 From chaos to statistical mechanics . . . 22
1.7 Semiclassical quantization . . . 23
1.8 Guide to literature . . . 25
Guide to exercises . . . 27
Resumé . . . 28
Exercises . . . 32

2 Flows **33**
2.1 Dynamical systems . . . 33
2.2 Flows . . . 37
2.3 Changing coordinates . . . 41
2.4 Computing trajectories . . . 44
2.5 Infinite-dimensional flows . . . 45
Resumé . . . 50
Exercises . . . 52

3 Maps **57**
3.1 Poincaré sections . . . 57
3.2 Constructing a Poincaré section . . . 60
3.3 Hénon map . . . 62
3.4 Billiards . . . 64
Exercises . . . 69

4 Local stability **73**
4.1 Flows transport neighborhoods . . . 73
4.2 Linear flows . . . 75
4.3 Nonlinear flows . . . 80
4.4 Hamiltonian flows . . . 82

4.5 Billiards . . . 83
4.6 Maps . . . 86
4.7 Cycle stabilities are metric invariants . . . 87
4.8 Going global: Stable/unstable manifolds . . . 91
Resumé . . . 92
Exercises . . . 94

5 Transporting densities **97**
5.1 Measures . . . 97
5.2 Density evolution . . . 99
5.3 Invariant measures . . . 102
5.4 Koopman, Perron-Frobenius operators . . . 105
Resumé . . . 110
Exercises . . . 112

6 Averaging **117**
6.1 Dynamical averaging . . . 117
6.2 Evolution operators . . . 124
6.3 Lyapunov exponents . . . 126
Resumé . . . 131
Exercises . . . 132

7 Trace formulas **135**
7.1 Trace of an evolution operator . . . 135
7.2 An asymptotic trace formula . . . 142
Resumé . . . 145
Exercises . . . 146

8 Spectral determinants **147**
8.1 Spectral determinants for maps . . . 148
8.2 Spectral determinant for flows . . . 149
8.3 Dynamical zeta functions . . . 151
8.4 False zeros . . . 155
8.5 More examples of spectral determinants . . . 155
8.6 All too many eigenvalues? . . . 158
Resumé . . . 161
Exercises . . . 163

9 Why does it work? **169**
9.1 The simplest of spectral determinants: A single fixed point . . . 170
9.2 Analyticity of spectral determinants . . . 173
9.3 Hyperbolic maps . . . 181
9.4 Physics of eigenvalues and eigenfunctions . . . 185
9.5 Why not just run it on a computer? . . . 188
Resumé . . . 192
Exercises . . . 194

10 Qualitative dynamics **197**
10.1 Temporal ordering: Itineraries . . . 198
10.2 Symbolic dynamics, basic notions . . . 200
10.3 3-disk symbolic dynamics . . . 204
10.4 Spatial ordering of "stretch & fold" flows . . . 206
10.5 Unimodal map symbolic dynamics . . . 210
10.6 Spatial ordering: Symbol square . . . 215
10.7 Pruning . . . 220
10.8 Topological dynamics . . . 222
Resumé . . . 230
Exercises . . . 233

11 Counting **239**
11.1 Counting itineraries . . . 239
11.2 Topological trace formula . . . 241
11.3 Determinant of a graph . . . 243
11.4 Topological zeta function . . . 247
11.5 Counting cycles . . . 249
11.6 Infinite partitions . . . 252
11.7 Shadowing . . . 255
Resumé . . . 257
Exercises . . . 260

12 Fixed points, and how to get them **269**
12.1 One-dimensional mappings . . . 270
12.2 d-dimensional mappings . . . 274
12.3 Flows . . . 275
12.4 Periodic orbits as extremal orbits . . . 279
12.5 Stability of cycles for maps . . . 283
Exercises . . . 288

13 Cycle expansions **293**
13.1 Pseudocycles and shadowing . . . 293
13.2 Cycle formulas for dynamical averages . . . 301
13.3 Cycle expansions for finite alphabets . . . 304
13.4 Stability ordering of cycle expansions . . . 305
13.5 Dirichlet series . . . 308
Resumé . . . 311
Exercises . . . 314

14 Why cycle? **319**
14.1 Escape rates . . . 319
14.2 Flow conservation sum rules . . . 323
14.3 Correlation functions . . . 325
14.4 Trace formulas *vs.* level sums . . . 326
Resumé . . . 329

Exercises . . . 331

15 Thermodynamic formalism **333**
15.1 Rényi entropies . . . 333
15.2 Fractal dimensions . . . 338
Resumé . . . 342
Exercises . . . 343

16 Intermittency **347**
16.1 Intermittency everywhere . . . 348
16.2 Intermittency for beginners . . . 352
16.3 General intermittent maps . . . 365
16.4 Probabilistic or BER zeta functions . . . 371
Resumé . . . 376
Exercises . . . 378

17 Discrete symmetries **381**
17.1 Preview . . . 382
17.2 Discrete symmetries . . . 386
17.3 Dynamics in the fundamental domain . . . 389
17.4 Factorizations of dynamical zeta functions . . . 393
17.5 C_2 factorization . . . 395
17.6 C_{3v} factorization: 3-disk game of pinball . . . 397
Resumé . . . 400
Exercises . . . 403

18 Deterministic diffusion **407**
18.1 Diffusion in periodic arrays . . . 408
18.2 Diffusion induced by chains of 1-d maps . . . 412
Resumé . . . 421
Exercises . . . 424

19 Irrationally winding **425**
19.1 Mode locking . . . 426
19.2 Local theory: "Golden mean" renormalization . . . 433
19.3 Global theory: Thermodynamic averaging . . . 435
19.4 Hausdorff dimension of irrational windings . . . 436
19.5 Thermodynamics of Farey tree: Farey model . . . 438
Resumé . . . 444
Exercises . . . 447

20 Statistical mechanics **449**
20.1 The thermodynamic limit . . . 449
20.2 Ising models . . . 452
20.3 Fisher droplet model . . . 455
20.4 Scaling functions . . . 461

20.5 Geometrization . . . 465
Resumé . . . 473
Exercises . . . 475

21 Semiclassical evolution **479**
21.1 Quantum mechanics: A brief review . . . 480
21.2 Semiclassical evolution . . . 484
21.3 Semiclassical propagator . . . 493
21.4 Semiclassical Green's function . . . 497
Resumé . . . 505
Exercises . . . 507

22 Semiclassical quantization **513**
22.1 Trace formula . . . 513
22.2 Semiclassical spectral determinant . . . 518
22.3 One-dimensional systems . . . 520
22.4 Two-dimensional systems . . . 522
Resumé . . . 522
Exercises . . . 527

23 Helium atom **529**
23.1 Classical dynamics of collinear helium . . . 530
23.2 Semiclassical quantization of collinear helium . . . 543
Resumé . . . 553
Exercises . . . 555

24 Diffraction distraction **557**
24.1 Quantum eavesdropping . . . 557
24.2 An application . . . 564
Resumé . . . 571
Exercises . . . 573

Summary and conclusions **575**
24.3 Cycles as the skeleton of chaos . . . 575

Index **580**

II Material available on `www.nbi.dk/ChaosBook/` **595**

A What reviewers say **597**
A.1 N. Bohr . . . 597
A.2 R.P. Feynman . . . 597
A.3 Divakar Viswanath . . . 597
A.4 Professor Gatto Nero . . . 597

B A brief history of chaos **599**
B.1 Chaos is born . . . 599
B.2 Chaos grows up . . . 603
B.3 Chaos with us . . . 604
B.4 Death of the Old Quantum Theory . . . 608

C Stability of Hamiltonian flows **611**
C.1 Symplectic invariance . . . 611
C.2 Monodromy matrix for Hamiltonian flows . . . 613

D Implementing evolution **617**
D.1 Material invariants . . . 617
D.2 Implementing evolution . . . 618
Exercises . . . 623

E Symbolic dynamics techniques **625**
E.1 Topological zeta functions for infinite subshifts . . . 625
E.2 Prime factorization for dynamical itineraries . . . 634

F Counting itineraries **639**
F.1 Counting curvatures . . . 639
Exercises . . . 641

G Applications **643**
G.1 Evolution operator for Lyapunov exponents . . . 643
G.2 Advection of vector fields by chaotic flows . . . 648
Exercises . . . 655

H Discrete symmetries **657**
H.1 Preliminaries and Definitions . . . 657
H.2 C_{4v} factorization . . . 662
H.3 C_{2v} factorization . . . 667
H.4 Symmetries of the symbol square . . . 670

I Convergence of spectral determinants **671**
I.1 Curvature expansions: geometric picture . . . 671
I.2 On importance of pruning . . . 675
I.3 Ma-the-matical caveats . . . 675
I.4 Estimate of the nth cumulant . . . 677

J Infinite dimensional operators **679**
J.1 Matrix-valued functions . . . 679
J.2 Trace class and Hilbert-Schmidt class . . . 681
J.3 Determinants of trace class operators . . . 683
J.4 Von Koch matrices . . . 687
J.5 Regularization . . . 689

K Solutions **693**

L Projects **723**
L.1 Deterministic diffusion, zig-zag map 725
L.2 Deterministic diffusion, sawtooth map 732

Viele Köche verderben den Brei

No man but a blockhead ever wrote except for money
Samuel Johnson

Predrag Cvitanović

most of the text

Roberto Artuso

5 Transporting densities 97
7.1.4 A trace formula for flows 140
14.3 Correlation functions 325
16 Intermittency 347
18 Deterministic diffusion 407
19 Irrationally winding 425

Ronnie Mainieri

2 Flows 33
3.2 The Poincaré section of a flow 60
4 Local stability 73
2.3.2 Understanding flows 43
10.1 Temporal ordering: itineraries 198
20 Statistical mechanics 449
Appendix B: A brief history of chaos 599

Gábor Vattay

15 Thermodynamic formalism 333
?? Semiclassical evolution ..... ??
22 Semiclassical trace formula 513

Ofer Biham

12.4.1 Relaxation for cyclists 280

Freddy Christiansen

12 Fixed points, and what to do about them 269

Per Dahlqvist

12.4.2 Orbit length extremization method for billiards 282
16 Intermittency 347

Appendix E.1.1: Periodic points of unimodal maps 631

Carl P. Dettmann

13.4 Stability ordering of cycle expansions 305

Mitchell J. Feigenbaum

Appendix C.1: Symplectic invariance 611

Kai T. Hansen

10.5 Unimodal map symbolic dynamics 210
10.5.2 Kneading theory 213
?? Topological zeta function for an infinite partition ..... **??**
figures throughout the text

Yueheng Lan

figures in chapters 1, and 17

Joachim Mathiesen

6.3 Lyapunov exponents 126
Rössler system figures, cycles in chapters 2, 3, 4 and 12

Adam Prügel-Bennet

Solutions 13.2, 8.1, 1.2, 3.7, 12.9, 2.11, 9.3

Lamberto Rondoni

5 Transporting densities 97
14.1.2 Unstable periodic orbits are dense 323

Juri Rolf

Solution 9.3

Per E. Rosenqvist

exercises, figures throughout the text

Hans Henrik Rugh

9 Why does it work? 169

Gábor Simon

Rössler system figures, cycles in chapters 2, 3, 4 and 12

Edward A. Spiegel

2 Flows 33
5 Transporting densities 97

Gregor Tanner

I.3 Ma-the-matical caveats 675
?? Semiclassical evolution ..... **??**
22 Semiclassical trace formula 513
23 The helium atom 529
Appendix C.2: Jacobians of Hamiltonian flows 613

Niall Whelan

24 Diffraction distraction 557
??: Trace of the scattering matrix ..... **??**

Andreas Wirzba

?? Semiclassical chaotic scattering ..... **??**
Appendix J: Infinite dimensional operators 679

Unsung Heroes: too numerous to list.

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 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 is a quick par-course of the main topics covered in the book. 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 B, and head straight to the starting line: A pinball game is used to motivate and illustrate most of the concepts to be developed in this book.

Throughout the book

indicates that the section 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

Learned remarks and bibliographical pointers are relegated to the "Commentary" section at the end of each chapter

1.1 Why this book?

> 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 1960's the pieces were counted, and in 1970's they were weighted and assembled together 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.

Then 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 1980's have overshadowed the deep insights of the 1970's, and these pictures have now migrated into textbooks. 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 and time scales dominated by chaotic dynamics.

Even though the experimental evidence for the fractal geometry of nature is circumstantial, in studies of probabilistically assembled fractal aggregates we know of nothing better than contemplating such quantities. In deterministic systems we can do *much* better. Chaotic dynamics is generated by interplay of locally unstable motions, and 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 the topology of chaotic dynamics is understood, a powerful theory yields the macroscopically measurable consequences of chaotic dynamics, such as atomic spectra, transport coefficients, gas pressures.

That is what we will focus on in this book. We teach you how to evaluate a determinant, take a logarithm, stuff like that. 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 pages subset of this text. Nothing to be done about that.

1.2 Chaos ahead

> Things fall apart; the centre cannot hold
> W.B. Yeats: *The Second Coming*

Study of chaotic dynamical systems 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 seem 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 for many by the recently acquired ability to numerically scan the phase space of non-integrable dynamical systems. The initial impression might be that all our analytic tools have failed us, and that the chaotic systems are amenable only to numerical and statistical investigations. However, as we show here, we already possess a theory of the deterministic chaos of predictive quality comparable to that of the traditional perturbation expansions for nearly integrable systems.

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; the asymptotic time phase space 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 tree of *cycles* (periodic orbits) of increasing lengths and self-similar structure. 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 linear stretching and folding of a baker's map, rather than the winding of a harmonic oscillator.

So, what is chaos, and what is to be done about it? To get some feeling for

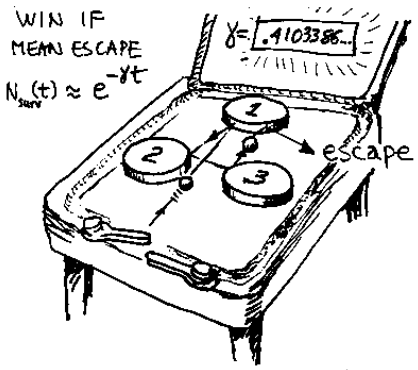


Figure 1.1: Physicists' bare bones game of pinball.

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 this book. 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.

1.3 A game of pinball

> Man må begrænse sig, det er en Hovedbetingelse for al Nydelse.
> Søren Kierkegaard, *Forførerens Dagbog*

That deterministic dynamics leads to chaos is no surprise to anyone who has tried pool, billiards or snooker – that is what the game is about – 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 diffusion constants in deterministic gases, or computing 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, fig. 1.1. Physicists' 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 18th century Baron Gottfried Wilhelm Leibniz was confident that given the initial conditions one knew what a deterministic system

would do far into the future. He wrote [1]:

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

Leibniz chose to illustrate his faith in determinism precisely with the type of physical system that we shall use here as a paradigm of "chaos". His claim is wrong in a deep and subtle way: a state of a physical system can *never* be specified to infinite precision, 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 fully determined by its initial conditions, in contra-distinction to a stochastic system, for which the initial conditions determine the present state only partially, due to noise, or other external circumstances beyond our control. For a stochastic system, 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 market to palpitations of chicken hearts. So, what is "chaos"?

Two pinball 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, fig. 1.2. This property of *sensitivity to initial conditions* can be quantified as

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

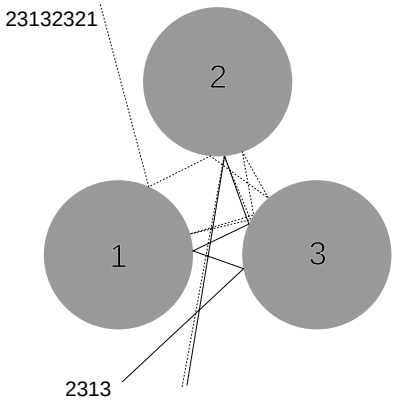


Figure 1.2: Sensitivity to initial conditions: two pinballs that start out very close to each other separate exponentially with time.

where λ, the mean rate of separation of trajectories of the system, is called the *Lyapunov exponent*. For any finite accuracy δx of the initial data, the dynamics is predictable only up to a finite *Lyapunov time*

☞ sect. 6.3

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

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 phase space and thus of necessity the separated trajectories are folded back and can re-approach each other arbitrarily closely, infinitely many times. In 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

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

☞ sect. 11.1

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

☞ sect. 15.1

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), 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 powerful is the Poincaré's vision of chaos as 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 phase space. Once this is grasped, the focus of theory shifts from attempting precise prediction of individual trajectories (which is impossible) to description of the geometry of the space of possible outcomes, and evaluation of averages over this space. How this is accomplished is what this book is about.

A definition of "turbulence" is harder to come by. Intuitively, the word refers to irregular behavior of an infinite-dimensional dynamical system (say, a bucket of boiling water) described by deterministic equations of motion (say, the Navier-Stokes equations). But in practice "turbulence" is very much like "cancer" - it is used 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.

☞ sect. 2.5

Confronted with a potentially chaotic dynamical system, we analyze it through a sequence of three distinct stages; diagnose, count, measure. I. First we determine the intrinsic *dimension* of the system – the minimum number of degrees of freedom necessary to capture its essential dynamics. If the system is very turbulent (description of its long time dynamics requires a space of high intrinsic dimension) we are, at present, out of luck. We know only how to deal with the transitional regime between regular motions and a few chaotic degrees of freedom. That is still something; even an infinite-dimensional system such as a burning flame front can turn out to have a very few chaotic degrees of freedom. In this regime the chaotic dynamics is restricted to a space of low dimension, the number of relevant parameters is small, and we can proceed to step II; we *count* and *classify* all possible topologically distinct trajectories of the system into a hierarchy whose successive layers require increased precision and patience on the part of the observer. This we shall do in sects. 1.3.3 and 1.3.4. If successful, we can proceed with step III of sect. 1.4.1: investigate the *weights* of the different pieces of the system.

☞ sect. 2.5
☞ chapter ??
☞ chapter 11

1.3.2 When does "chaos" matter?

> Whether 'tis nobler in the mind to suffer
> The slings and arrows of outrageous fortune,
> Or to take arms against a sea of troubles,
> And by opposing end them?
> W. Shakespeare, *Hamlet*

When should we be mindfull 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 in which phase space regions comparable in size to the observational accuracy extend across the entire accessible phase space, is significantly shorter than the observational time, we need methods that will be developped 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.

As in science popularizations too much has been made of the impact of the "chaos theory" , perhaps it is not amiss to state a number of caveats already at this point.

At present the theory is in practice applicable only to systems with a low intrinsic *dimension* – the minimum number of degrees of freedom necessary to capture its essential dynamics. If the system is very turbulent (description of its long time dynamics requires a space of high intrinsic dimension) we are out of luck. Hence insights that the theory offers to elucidation of 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.
☞ sect. 2.5
There are applications – such as spatially extended systems and statistical mechanics applications – where the few important degrees of freedom can be isolated
☞ chapter 18
and studied profitably by methods to be described here.

The theory has had limited practical success applied to the very noisy systems so important in 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 interburst intervals, cardiac pulse, etc.), disentangling "chaotic" motions from the environmental noise has been very hard.

1.3.3 Symbolic dynamics

> Formulas hamper the understanding.
> S. Smale

We commence our analysis of the pinball game with steps I, II: diagnose, count. We shall return to step III – measure – in sect. 1.4.1.
☞ chapter 13

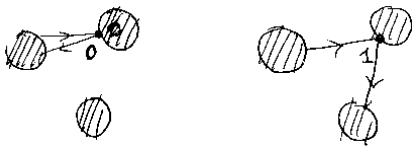


Figure 1.3: Binary labeling of the 3-disk pinball trajectories; a bounce in which the trajectory returns to the preceding disk is labeled 0, and a bounce which results in continuation to the third disk is labeled 1.

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 which indicates the order in which the disks are visited; for example, the two trajectories in fig. 1.2 have itineraries _2313_, _23132321_ respectively. The itinerary will be 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. Parenthetically, in this subject the words "orbit" and "trajectory" refer to one and the same thing.

1.1 on p. 32

Such labeling is the simplest example of *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 - there are no further pruning rules.

chapter ??

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, fig. 1.3. A clever choice of an alphabet will incorporate important features of the dynamics, such as its symmetries.

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.

sect. 24.3

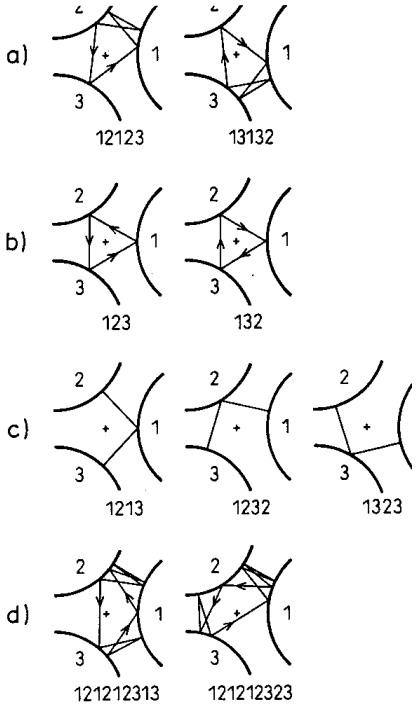


Figure 1.4: Some examples of 3-disk cycles: (a) $\overline{12123}$ and $\overline{13132}$ are mapped into each other by σ_{23}, the flip across 1 axis; this cycle has degeneracy 6 under C_{3v} symmetries. (C_{3v} is the symmetry group of the equilateral triangle.) Similarly (b) $\overline{123}$ and $\overline{132}$ and (c) $\overline{1213}$, $\overline{1232}$ and $\overline{1323}$ are degenerate under C_{3v}. (d) The cycles $\overline{121212313}$ and $\overline{121212323}$ are related by time reversal but not by any C_{3v} symmetry. These symmetries are discussed in more detail in chapter 17. (from ref. [2])

1.3.4 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 - some examples are drawn in fig. 1.4 - but it is rather hard to perceive the systematics of orbits from their shapes. In the pinball example the problem is that we are looking at the projections of a 4-dimensional phase space trajectories onto a 2-dimensional subspace, the space coordinates. While the trajectories cannot intersect (that would violate their deterministic uniqueness), their projections on arbitrary subspaces intersect in a rather arbitrary fashion. A clearer picture of the dynamics is obtained by constructing a phase space Poincaré section.

The position of the ball is described by a pair of numbers (the spatial coordinates on the plane) and its velocity by another pair of numbers (the components of the velocity vector). As far as baron Leibniz is concerned, this is a complete description.

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 four-dimensional flow to a two-dimensional map f that takes the coordinates of the pinball from one disk edge to another disk edge.

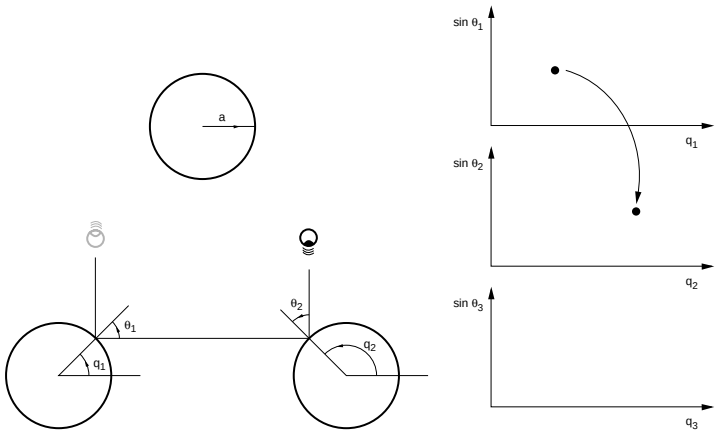
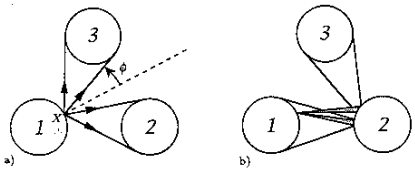


Figure 1.5: (a) The 3-disk game of pinball coordinates and (b) the Poincaré sections.

Figure 1.6: (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. The pencils of initial conditions that hit more and more consecutive disks are nested within each other as in fig. 1.7.



Let us state this more precisely: the trajectory just after the moment of impact is defined by marking q_i, the arc-length position of the ith bounce along the billiard wall, and $p_i = \sin\theta_i$, the momentum component parallel to the billiard wall at the point of impact, fig. 1.5. Such section of a flow is called a *Poincaré section*, and the particular choice of coordinates (due to Birkhoff) is particulary smart, as it conserves the phase-space volume. In terms of the Poincaré section, the dynamics is reduced to the *return map* $f : (p_i, q_i) \mapsto (p_{i+1}, q_{i+1})$ from the boundary of a disk to the boundary of the next disk. The explicit form of this map is easily written down, but it is of no importance right now.

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 $\mathcal{M}_0$, $\mathcal{M}_1$. Embedded within them there are four strips $\mathcal{M}_{00}$, $\mathcal{M}_{10}$, $\mathcal{M}_{01}$, $\mathcal{M}_{11}$ of initial conditions that survive for two bounces, and so forth, see figs. 1.6 and 1.7. Provided that the disks are sufficiently separated, after n bounces the survivors are divided into 2^n distinct strips: the ith strip consists of all points with itinerary $i = s_1 s_2 s_3 \ldots s_n$, $s = \{0, 1\}$. 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 have a symbolic dynamics; it provides a navigation

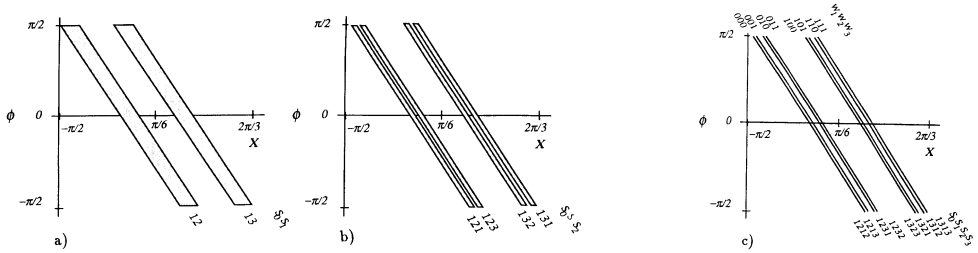


Figure 1.7: Ternary labelled regions of the 3-disk game of pinball phase space Poincaré section which correspond to trajectories that originate on disk 1 and remain confined for (a) one bounce, (b) two bounces, (c) three bounces. The Poincaré sections for trajectories originating on the other two disks are obtained by the appropriate relabelling of the strips (K.T. Hansen [3]).

chart through chaotic phase 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.

1.3.5 Escape rate

What is a good physical quantity to compute for the game of pinball? A repeller *escape rate* is an eminently measurable quantity. An example of such measurement would be an unstable molecular or nuclear state which can be well approximated by a classical potential with possibility of escape in certain directions. In an experiment many projectiles are injected into such a non-confining potential and their mean escape rate is measured, as in fig. 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.

1.2 on p. 32

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 the 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 fig. 1.7 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

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

where i is a label of the ith strip, $|\mathcal{M}|$ is the initial area, and $|\mathcal{M}_i|$ is the area of the ith strip of survivors. 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

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

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

1.4 Periodic orbit theory

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

If, when asked what the 3-disk escape rate is for disk radius 1, center-center separation 6, velocity 1, you answer that the continuous time escape rate is roughly $\gamma = 0.410338407769346489338461307819 2\ldots$, you do not need this book. If you have no clue, hang on.

1.4.1 Size of a partition

Not only do the periodic points keep track of locations and the ordering of the strips, but, as we shall now show, they also determine their size.

As a trajectory evolves, it carries along and distorts its infinitesimal neighborhood. Let

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

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

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

☞ sect. 4.5 Evaluation of a cycle Jacobian matrix is a longish exercise - here we just state the result. The Jacobian matrix describes the deformation of an infinitesimal neighborhood of $x(t)$ as it goes with the flow; its the eigenvectors and eigenvalues give the directions and the corresponding rates of its expansion or contraction. The trajectories that start out in an infinitesimal neighborhood are separated along the unstable directions (those whose eigenvalues are less than unity in magnitude), approach each other along the stable directions (those whose eigenvalues exceed unity in magnitude), and maintain their distance along the marginal directions (those whose eigenvalues equal unity in magnitude). In our game of pinball after one traversal of the cycle p the beam of neighboring trajectories is defocused in the unstable eigendirection by the factor Λ_p, the expanding eigenvalue of the 2-dimensional surface of section return map Jacobian matrix $\mathbf{J}_p$.

As the heights of the strips in fig. 1.7 are effectively constant, we can concentrate on their thickness. If the height is $\approx L$, then the area of the ith strip is $\mathcal{M}_i \approx L l_i$ for a strip of width l_i.

Each strip i in fig. 1.7 contains a periodic point x_i. The finer the intervals, the smaller is the variation in flow across them, and 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 Λ_i is the unstable eigenvalue of the i'th periodic point (due to the low dimensionality, the Jacobian can have at most one unstable eigenvalue.) Note that it is the magnitude of this eigenvalue which is important and 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 them to ☞ sect. 5.3 be insensitive to smooth variations in the initial distribution.

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 ☞ sect. 7.1.1 of Λ_i, so we neglect them. If the hyperbolicity assumption is justified, we can

replace $|\mathcal{M}_i| \approx L l_i$ in (1.2) by $1/|\Lambda_i|$ and consider the sum

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

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

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

This is the property of $\Gamma(z)$ which motivated its definition. We now devise an alternate expression for (1.5) in terms of periodic orbits to make explicit the connection between the escape rate and the 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 γ is now given by the leading pole of (1.7), rather than a numerical extrapolation of a sequence of γ_n extracted from (1.3). ☞ sect. 7.2

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.

1.4.2 Dynamical zeta function

If a trajectory retraces a *prime* cycle r times, its expanding eigenvalue is Λ_p^r. A prime cycle p is a single traversal of the orbit; its label is a non-repeating symbol string of n_p symbols. There is only one prime cycle for each cyclic permutation class. For example, $p = \overline{0011} = \overline{1001} = \overline{1100} = \overline{0110}$ 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

11.5 on p. 261

sect. 4.6

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

where the index p runs through all distinct *prime* cycles. Note that we have resumed the contribution of the cycle p to all times, so truncating the summation up to given p is *not* a finite time $n \leq n_p$ approximation, but an asymptotic, *infinite* time estimate based by approximating stabilities of all cycles by a finite number of the shortest cycles and their repeats. The $n_p z^{n_p}$ factors in (1.8) 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 choice of "zeta" in its definition as $1/\zeta(z)$. This is the prototype formula of the 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.

sect. 14.1

1.4.3 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's method searches for periodic solutions; we shall assume that the numerics is 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

☞ sect. 12.4.2

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 13

☞ sect. 13.1

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 γ by finding the smallest $z = e^{\gamma}$ for which (1.10) vanishes.

1.4.4 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?

☞ sect. 13.1.3

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 fig. 1.8; 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

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 fig. 1.3, the trajectory is alternating between 0 and 1. The incoming and outgoing angles when it executes these bounces are very close to the corresponding angles for 0 and 1 cycles. Also the distances traversed between bounces are similar so that the 2-cycle expanding eigenvalue Λ_{01} is close in magnitude to the product of the 1-cycle eigenvalues $\Lambda_0 \Lambda_1$.

To understand this on a more general level, try to visualize the partition of a chaotic dynamical system's phase space in terms of cycle neighborhoods as a tessellation of the dynamical system, with smooth flow approximated by its periodic orbit skeleton, each "face" centered on a periodic point, and the scale of the "face" determined by the linearization of the flow around the periodic point, fig. 1.8.

The orbits that follow the same symbolic dynamics, such as $\{ab\}$ and a "pseudo orbit" $\{a\}\{b\}$, lie close to each other in the phase 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 9

1.5 Evolution operators

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

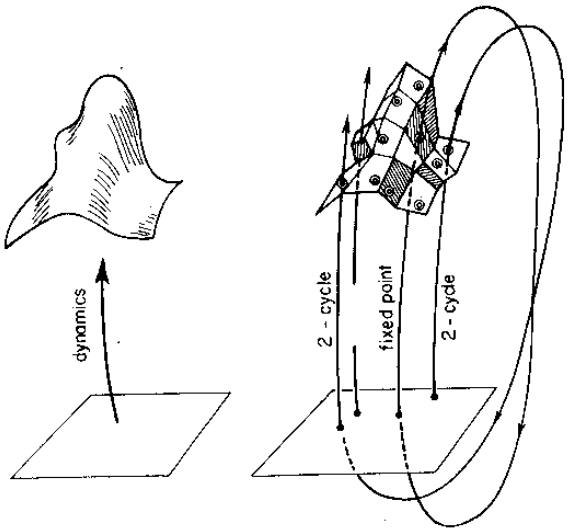


Figure 1.8: Approximation to (a) a smooth dynamics by (b) the skeleton of periodic points, together with their linearized neighborhoods. Indicated are segments of two 1-cycles and a 2-cycle that alternates between the neighborhoods of the two 1-cycles, shadowing first one of the two 1-cycles, and then the other.

rate measured in units of continuous time. For continuous time flows, the escape rate (1.2) is generalized as follows. Define a finite phase 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 fig. 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

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

is the Dirac delta function, as for a deterministic flow the initial point y maps into a unique point x at time t. For discrete time, $f^n(x)$ is the nth 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

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

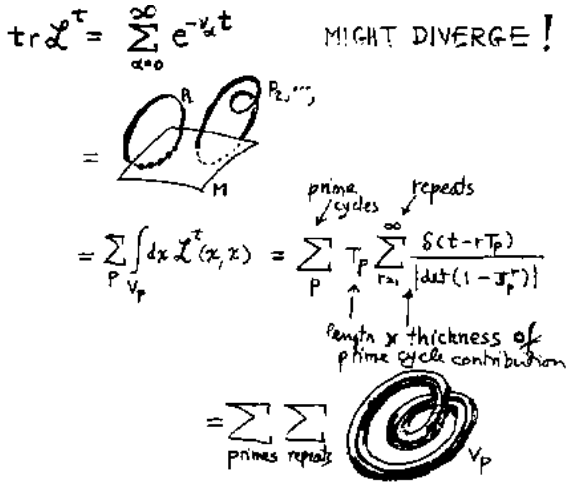


Figure 1.9: The trace of an evolution operator is concentrated in tubes around prime cycles, of length T_p and thickness $1/|\Lambda_p|^r$ for rth repeat of the prime cycle p.

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 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 chaotic dynamics might reveal it to be complex beyond belief, it is still generated by a simple deterministic law, and with some luck and insight, our labeling of possible motions will reflect this simplicity. If the rule that gets us from one level of the classification hierarchy to the next does not depend strongly on the level, the resulting hierarchy is approximately self-similar. We now turn such approximate self-similarity to our advantage, by turning it into an operation, the action of the evolution operator, whose iteration encodes the self-similarity.

1.5.1 Trace formula

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 the *trace formulas* and the *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 phase space. The result is an expression for $\mathrm{tr}\,\mathcal{L}^t$ as a sum over neighborhoods of prime cycles p and their repetitions

☞ sect. 7.1.4

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

This formula has a simple geometrical interpretation sketched in fig. 1.9. After the rth 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 $1/\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right| \to 1/|\Lambda_p|$.

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 $\mathrm{tr}\,e^{\mathcal{A}} = \sum e^{s_\alpha t}$, while the right-hand side equals zero everywhere except for the set $t = rT_p$. A Laplace transform smoothes 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:

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

The beauty of the 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 $\mathbf{J}_p$ – is an invariant property of the flow, independent of any coordinate choice.

☞ sect. 7.1

1.5.2 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

1.3 on p. 32

$$\begin{aligned}\ln\det(s-\mathcal{A}) &= \mathrm{tr}\,\ln(s-\mathcal{A}) \\ \frac{d}{ds}\ln\det(s-\mathcal{A}) &= \mathrm{tr}\,\frac{1}{s-\mathcal{A}}\,,\end{aligned}$$

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 8

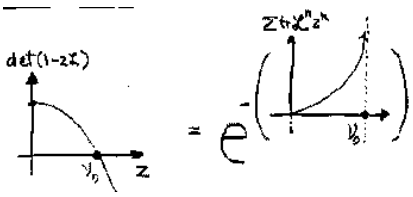


Figure 1.10: Spectral determinant is preferable to the trace as it vanishes smoothly at the leading eigenvalue, while the trace formula diverges.

$$\det(s-\mathcal{A}) = \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{-sT_p r}}{\left|\det\left(\mathbf{1}-\mathbf{J}_p^r\right)\right|}\right) . \tag{1.15}$$

The s integration leads here to replacement $T_p \to T_p/rT_p$ in the periodic orbit expansion (1.14).

The motivation for recasting the eigenvalue problem in this form is sketched in fig. 1.10; exponentiation improves analyticity and trades in a divergence of the trace sum for a zero of the spectral determinant. The computation of the zeros of $\det(s-\mathcal{A})$ proceeds very much like the computations of sect. 1.4.3.

☞ sect. 8.5.1

1.6 From chaos to statistical mechanics

While the above replacement of dynamics of individual trajectories by evolution operators which propagate densities might feel like just another bit of mathematical voodoo, actually something very radical has taken place. Consider a chaotic flow, such as 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 stirring, we would return back to the perfect white/red separation. However, we know that this cannot be true – 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.

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, but instead gain 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 (such as the *stosszahlansatz* of

☞ chapter 18

Boltzmann).

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 the equilibrium statistical mechanics.

The results of the 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 16
☞ sect. ??

1.7 Semiclassical quantization

So far, so good – anyone can play a game of classical pinball, and a skilled neuroscientist can poke rat brains. But what happens quantum mechanically, that is, if we scatter waves rather than point-like pinballs? Were the game of pinball a closed system, quantum mechanically one would determine its stationary eigenfunctions and eigenenergies. For open systems one seeks instead for complex resonances, where the imaginary part of the eigenenergy describes the rate at which the quantum wave function leaks out of the central multiple scattering region. One of the pleasant surprises in the development of the theory of chaotic dynamical systems was the discovery that the zeros of dynamical zeta function (1.9) also yield excellent estimates of *quantum* resonances, with the quantum amplitude associated with a given cycle approximated semiclassically by the "square root" of the classical weight (1.15)

$$t_p = \frac{1}{\sqrt{|\Lambda_p|}} e^{\frac{i}{\hbar} S_p - i\pi m_p/2} \ . \tag{1.16}$$

Here the phase is given by the Bohr-Sommerfeld action integral S_p, together with an additional topological phase m_p, the number of points on the periodic trajectory where the naive semiclassical approximation fails us. ☞ chapter ??

1.7.1 Quantization of helium

Now we are finally in position to accomplish something altogether remarkable; we put together all ingredients that made the pinball unpredictable, and compute a "chaotic" part of the helium spectrum to shocking accuracy. Poincaré

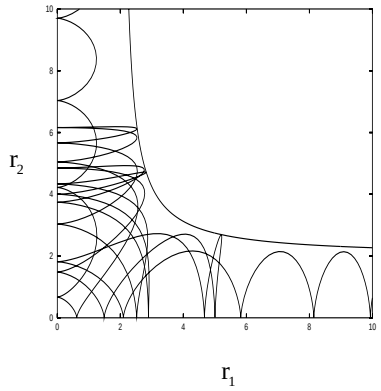


Figure 1.11: A typical collinear helium trajectory in the $r_1 - r_2$ plane; the trajectory enters along the r_1 axis and escapes to infinity along the r_2 axis.

taught us that from the classical dynamics point of view, helium is an example of the 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, and the nucleus mass to ∞. In these units the helium nucleus has charge 2, the electrons have charge -1, and the Hamiltonian is

$$H = \frac{1}{2}p_1^2 + \frac{1}{2}p_2^2 - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_1 + r_2}\,. \tag{1.17}$$

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, or, better still, by an appropriately chosen 2-d 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 again 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* (1.16) 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 23

1.8 Guide to literature

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

This text aims to bridge the gap between the physics and mathematics dynamical systems literature. The intended audience is the dream graduate student, with a theoretical bent. As a complementary presentation we recommend Gaspard's monograph [4] which covers much of the same ground in a highly readable and scholarly manner.

As far as the prerequisites are concerned - this book 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 [5], an introduction to flows, fixed points, manifolds, bifurcations. It is probably the most accessible introduction to nonlinear dynamics - it starts out with differential equations, and its broadly chosen examples and many exercises make it favorite with students. It is not strong on chaos. There the textbook of Alligood, Sauer and Yorke [6] is preferable: an elegant introduction to maps, chaos, period doubling, symbolic dynamics, fractals, dimensions - a good companion to this book. An introduction more comfortable to physicists is the textbook by Ott [7], with baker's map used to illustrate many key techniques in analysis of chaotic systems. It is perhaps harder than the above two as the first book on nonlinear dynamics.

The 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 dynamics. With this, and 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 this book. The courses we have taught start out with the introductory chapters on qualitative dynamics, symbolic dynamics and flows, and than 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.

Spatiotemporal dynamical systems. Partial differential equations for dissipative systems, weak amplitude expansions, normal forms, symmetries and bifurcations, pseudospectral methods, spatiotemporal chaos.

Quantum chaology. Semiclassical propagators, density of states, trace for-

mulas, semiclassical spectral determinants, billiards, semiclassical helium, diffraction, creeping, tunneling, higher $\hbar$ corrections.

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. The role of unstable periodic orbits was already fully appreciated by Poincaré [8, 9], 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 [10] for an overview of some of that literature.

If you find this book not rigorous enough, you should turn to the mathematics literature. The most extensive reference is the treatise by Katok and Hasselblatt [11], an impressive compendium of modern dynamical systems theory. The fundamental papers in this field, all still valuable reading, are Smale [12], Bowen [13] and Sinai [14]. 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[?]. Markov partitions were introduced by Sinai in ref. [15]. The classical text (though certainly not an easy read) on the subject of dynamical zeta functions is Ruelle's *Statistical Mechanics, Thermodynamic Formalism* [16]. 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 [18]. Further excellent mathematical references on thermodynamic formalism are Parry and Pollicott's monograph [19] with emphasis on the symbolic dynamics aspects of the formalism, and Baladi's clear and compact reviews of dynamical zeta functions [20, 21].

A graduate level introduction to statistical mechanics from the dynamical point view is given by Dorfman [22]; the Gaspard monograph [4] covers the same ground in more depth. Driebe monograph [23] 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?"* [24].

A key prerequisite to developing any theory of "quantum chaos" is solid understanding of the Hamiltonian mechanics. For that, Arnold's text [25] is the essential reference. Ozorio de Almeida [26] is a nice introduction of the aspects of Hamiltonian dynamics prerequisite to 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 [27] is an excel-

lent introduction to the semiclassical methods. Gutzwiller's monograph [28] 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 chaology. Perhaps more suitable as a graduate course text is Reichl's presentation [29]. For an introduction to "quantum chaos" that focuses on the random matrix theory the reader can consult the monograph by Haake [30], among others.

If you were wandering while reading this introduction "what's up with rat brains?", the answer is yes indeed, there is a line of research in study on neuronal dynamics that focuses on possible unstable periodic states, described for example in ref. [31].

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. Some of the solutions provided might be more illuminating than the main text. So 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. The problems that you should do have **underlined titles**. *The rest (smaller type) are optional.* Difficult optional problems are marked by any number of *** stars. By the end of the course you should have completed at least three projects: (a) compute everything for a one-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:

13.2 on p. 314

- **Dynamics**
  1. count prime cycles, exercise 1.1, exercise 10.1, exercise 10.4
  2. pinball simulator, exercise 3.7, exercise 12.4
  3. pinball stability, exercise 4.4, exercise 12.4
  4. pinball periodic orbits, exercise 12.5, exercise 12.6
  5. helium integrator, exercise 2.11, exercise 12.7
  6. helium periodic orbits, exercise 23.4, exercise 12.8

- **Averaging, numerical**
  1. pinball escape rate, exercise 8.11
  2. Lyapunov exponent, exercise 15.2
- **Averaging, periodic orbits**
  1. cycle expansions, exercise 13.1, exercise 13.2
  2. pinball escape rate, exercise 13.4, exercise 13.5
  3. cycle expansions for averages, exercise 13.1, exercise 14.3
  4. cycle expansions for diffusion, exercise 18.1
  5. pruning, Markov graphs
  6. desymmetrization exercise 17.1
  7. intermittency, phase transitions
  8. semiclassical quantization exercise 22.4
  9. ortho-, para-helium, lowest eigenenergies exercise 23.7

Solutions for some of the problems are included appendix K. Often going through a solution is more instructive than reading the corresponding chapter.

Résumé

The goal of 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 phase space. Hence the theory focuses on description of the geometry of the space of possible outcomes, and evaluation of averages over this space, rather than attempting the impossible, precise prediction of individual trajectories. The dynamics of distributions 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 the *trace formulas* and the *spectral determinants*.

The theory of evaluation of 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 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 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 then an efficient method for evaluating classical and quantum observables.

The critical step in the derivation of the dynamical zeta function was the hyperbolicity assumption, that is 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 exposition simple we have here illustrated the utility of cycles and their curvatures by a pinball game, but topics covered in this book – 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 and zeta functions, and so on to the semiclassical quantization of helium – should give the reader some confidence in the general applicability 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. One of the surprises is that the quantum mechanics of classically chaotic systems is very much like the classical mechanics of chaotic systems; both are described by nearly the same zeta functions and cycle expansions, with the same dependence on the topology of the classical flow.

References

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

[1.2] P. Cvitanović, B. Eckhardt, P.E. Rosenqvist, G. Russberg and P. Scherer, in G. Casati and B. Chirikov, eds., *Quantum Chaos* (Cambridge University Press, Cambridge 1993).

[1.3] K.T. Hansen, *Symbolic Dynamics in Chaotic Systems*, Ph.D. thesis (Univ. of Oslo, 1994).
`www.nbi.dk/CATS/papers/khansen/thesis/thesis.html`

[1.4] P. Gaspard, *Chaos, Scattering and Statistical Mechanics* (Cambridge Univ. Press, Cambridge 1997).

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

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

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

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

[1.9] 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).

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

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

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

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

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

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

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

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

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

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

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

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

[1.22] R. Dorfman, *From Molecular Chaos to Dynamical Chaos* (Cambridge Univ. Press, Cambridge 1998).

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

[1.24] J. Bricmont, *"Science of Chaos or Chaos in Science?"*, available on `www.ma.utexas.edu/mp_arc, #96-116.`

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

[1.26] A.M. Ozorio de Almeida, *Hamiltonian Systems: Chaos and Quantization* (Cambridge University Press, Cambridge, 1988).

[1.27] M. Brack and R.K. Bhaduri, *Semiclassical Physics* (Addison-Wesley, New York 1997).

[1.28] M.C. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer, New York 1990).

[1.29] L.E. Reichl, *The Transition to Chaos in Conservative Classical Systems: Quantum Manifestations* (Springer-Verlag, New York, 1992).

[1.30] F. Haake, *Quantum Signatures of Chaos* (Springer-Verlag, New York, 1991).

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

Exercises

1.1 3-disk symbolic dynamics. As the periodic trajectories will turn out to be the our main tool to breach deep into the realm of chaos, it pays to start familiarizing oneself with them already now, by sketching and counting the few shortest prime cycles (we return to this in sect. 11.4). Show that the 3-disk pinball has $3 \cdot 2^n$ 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 game without cheating, by hook or crook (by the end of chapter 13 you should be in position to make very accurate estimates).

1.3 Trace-log of a matrix. Prove that

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

for arbitrary finite dimensional matrix M.

Chapter 2

Flows

Poetry is what is lost in translation
Robert Frost

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

We start out by a recapitulation of the basic notions of dynamics. Our aim is narrow; keep the exposition focused on prerequsites to the applications to be developed in this text. We assume that the reader is familiar with the dynamics on the level of introductory texts mentioned in sect. 1.8, 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 anyway beyond human ken. 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 chapter 5.

fast track:
chapter 5, p. 97

2.1 Dynamical systems

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, the future behavior can be predicted. 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, the latitude and longitude in the celestial sphere are enough to completly specify the planet's motion. All possible values for positions and velocities of the planets form the *phase space* of the system. More generally, a state of a physical system at a given instant in time can be represented by a single point in an abstract space called *state space* or *phase space* $\mathcal{M}$. As the system changes, so does the *representative point* in phase 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 phase space and maps it into another point. Not two or three, but exactly one. This is not always possible. For example, knowing the temperature today is not enough to predict the temperature tommorrow; or knowing the value of a stock market index today will not determine its value tommorrow. The phase space can be enlarged, in the hope that in a sufficently large phase 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 tommorrow. Even that is not quite true, and we are less hopeful when it comes to a stock index.

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. ☞ chapter 10.6.1 We may think such sets a nuisance, but it is quite the contrary - will enable us to partition phase space so that the dynamics can be better understood.

Locally the phase space $\mathcal{M}$ is $\mathbb{R}^d$, 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 manifold. 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* or *dynamics* $f^t : \mathcal{M} \to \mathcal{M}$ that tells where a point x is in $\mathcal{M}$ after a time interval t. The pair $(\mathcal{M}, f)$ is called 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 $f(t, x)$ to remind us that f is really a function of two variables: time interval and point of phase space. Notice that time is not absolute, only the time interval is necessary. This is because a point in phase space completely determines all future evolution and it is not necessary to know anything else. The time parameter can be a real variable ($t \in \mathbb{R}$), in which case the evolution is called a *flow*, or an integer ($t \in \mathbb{Z}$), in which case the evolution advances in discrete steps in time, given by *iteration*

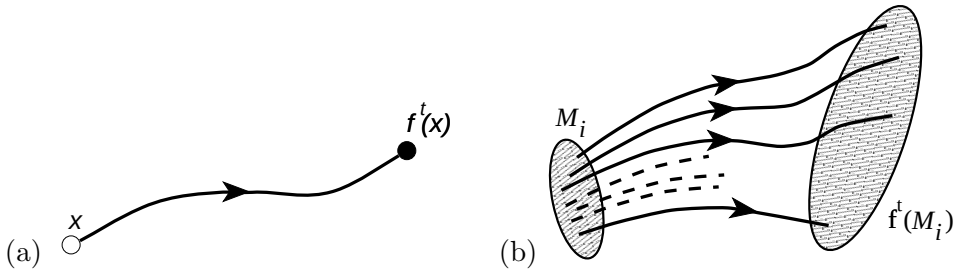


Figure 2.1: (a) A trajectory traced out by the evolution rule f^t. Starting from the phase space point x, after a time t, the point is at $f^t(x)$. (b) The evolution rule f^tcan be used to map a region $\mathcal{M}_i$ of the phase space into $f^t(\mathcal{M}_i)$.

of a *map*.

Nature provides us with inumerable 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 $x_0 = x(0)$. Because f^t is a single-valued function, any point of the trajectory can be used to label the trajectory. We can speak of the trajectory starting at x_0, or of the trajectory passing through a point $y = f^t(x_0)$. For flows the trajectory of a point is a continuous curve; for a map, a sequence of points. By extension, we can also talk of the evolution of a region $\mathcal{M}_i$ of the phase space: just apply f^t to every point in $\mathcal{M}_i$ to obtain a new region $f^t(\mathcal{M}_i)$, as in fig. 2.1.

2.1 on p. 52

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

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 $t \neq t'$.

The ancients no less than the contemporary field theorists tried to make sense of all dynamics in terms of periodic motions; epicycles, integrable systems. Embarassing truth is that for a generic dynamical systems most motions are aperiodic. We will break aperiodic motions up into two types: 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

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

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

A periodic trajectory is an example of a trajectory that returns exactly to the initial point in a finite time; however, periodic trajectories are a very small subset of the phase space, in the same sense that rationals are a set of zero measure on the unit interval. 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

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

In other words, the trajectory of a non-wandering point reenters the neighborhood $\mathcal{M}_0$ infinitely often. We shall denote by Ω the *non–wandering set* of f, that is the union of all the non-wandering points of $\mathcal{M}$. The set Ω, 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 phase space volume that maps into itself under the forward evolution (by the method of Lyapunov functionals, or any other method), the flow is globally contracting onto a subset of $\mathcal{M}$ that 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 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 reccurent attractor to which we shall refer loosely as a *strange attractor*. We say loosely, as it will soon become apparent that diagnosing and proving existence of a genuine, card carrying strange attractor is a tricky undertaking.

☞ sect. 2.2.1

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

It would seem that having said that the periodic points are too exceptional, and that almost all non-wandering points are aperiodic, we have given up the ancients' fixation on periodic motions. Not so. As longer and longer cycles approximate more and more accurately finite segments of aperiodic trajectories, we shall establish control over the non–wandering set 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 into a little more detail.

2.2 Flows

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:

2.2 on p. 52

(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.

The family of mappings $f^t(x)$ thus forms a continuous (forward 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 $f^{-t}(f^t(x)) = f^0(x) = x$, and thus 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 infinite dimensional systems, for systems with noise, for non-invertible maps - time reversal is not an option, 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.

sect. 2.5

For infinitesimal times flows can be defined by differential equations. Write a trajectory as

$$x(t+\tau) = f^{t+\tau}(x_0) = f(\tau, f(t, x_0)) \tag{2.3}$$

and compute the τ derivative

$$\left.\frac{dx}{d\tau}\right|_{\tau=0} = \left.\frac{\partial f}{\partial \tau}(\tau, f(t,x_0))\right|_{\tau=0} = \frac{\partial}{\partial t} f^0(x(t))\,. \tag{2.4}$$

2.3 on p. 52

$\dot{x}(t)$, the time derivative of a trajectory at point $x(t)$, can be expressed as the time derivative of the evolution rule, a vector evaluated at the same point. By considering all possible trajectories, we obtain the vector $\partial_t f^0(x)$ at any point $x \in \mathcal{M}$ and define a *vector field*

$$v(x) = \frac{\partial f^0}{\partial t}(x)\,. \tag{2.5}$$

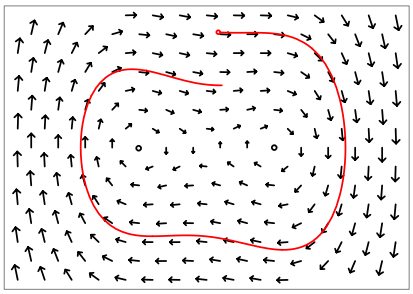
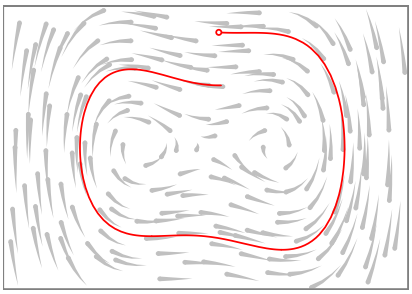
(a)  (b) 

Figure 2.2: (a) The two-dimensional vector field for the Duffing system (2.7), together with a short trajectory segment. The vectors are drawn superimposed over the configuration coordinates $(x(t), y(t))$ of phase space $\mathcal{M}$, but they belong to a different space, the *tangent bundle* $T\mathcal{M}$. (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.

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. An equation that is second or higher order in time can always be rewritten as a set of first order equations. Here we are concerned with a much larger world of general flows, mechanical or not, all defined by a time independent vector field

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

At each point of the phase space there is a vector that gives the direction in which the orbit will evolve. As a concrete example, consider the two-dimensional vector field for the 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 two ways in fig. 2.2. The length of the vector is proportional to the speed of the point, and its direction and length changes from point to point. When the phase space is a manifold more complicated than $\mathbb{R}^d$, one can no longer think of the vector field as being embedded in phase space. Instead, we have to imagine that each point x of phase space has a different tangent plane $T\mathcal{M}_x$ attached to it, and even if these planes seem to cross when they are drawn on a piece of paper, they do not. The vector field lives in the union of all these tangent planes, a space called the *tangent bundle* $T\mathcal{M}$.

2.4 on p. 52

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

x_q is an *equilibrium point* (often referred to as a *stationary*, *fixed*, or *stagnation* point) and the trajectory remains forever stuck at x_q. Otherwise the trajectory is 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 the *autonomous* or *stationary* flows, that is flows for which the velocity field v_i is not explicitely dependent on time. If you insist on studying a non-autonomous system

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

we can always convert it into a system where time does not appear explicitly. To do so, extend the phase space to $(d+1)$-dimensional $x = \{y, \tau\}$ and the vector field to

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

2.5 on p. 53

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.

2.2.1 A flow with a strange attractor

> There is no beauty without some strangeness
> William Blake

A concrete example of an autonomous flow is the *Rössler system*

$$\begin{aligned} \dot{x} &= -y - z \\ \dot{y} &= x + ay \\ \dot{z} &= b + z(x-c)\,, \qquad a = b = 0.2\,, \quad c = 5.7\,. \end{aligned} \tag{2.12}$$

The system is as simple as they get - it would be linear were it not for the sole quadratic term zx. Even for so simple a system, the nature of long-time solutions is far from obvious. Close to the origin there is a repelling equilibrium point, but to see what other solutions look like we need to resort to numerical integration.

A typical numerically integrated long-time trajectory is sketched in fig. 2.3. As we shall show in sect. 4.1, for this flow any finite volume of initial conditions

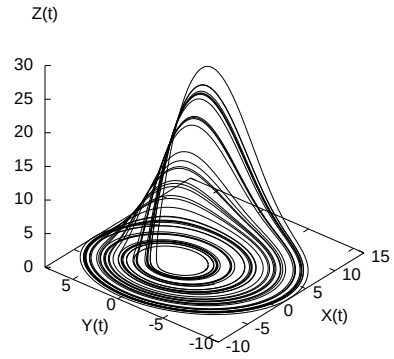


Figure 2.3: A trajectory of the Rössler flow at time $t = 250$. (G. Simon)

shrinks with time, so the flow is *contracting*. All trajectories seem to converge to a *strange attractor*. We say "seem", as there exist no proof that this attractor is strange. For now, accept that fig. 2.3 and similar figures in what follows are examples of "strange attractors".

You might think that this strangeness has to do with contracting flows only. Not at all - we chose this example as it is easier to visualise aperiodic dynamics when the flow is contracting onto a lower-dimensional attracting set. As the next example we take a flow that preserves phase space volumes.

2.2.2 A Hamiltonian flow

☞ appendix C

An important class of dynamical systems are the Hamiltonian flows, given by a time-independent Hamiltonian $H(q, p)$ together with the Hamilton's equations of motion

$$\dot{q}_i = \frac{\partial H}{\partial p_i}\,, \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i}\,, \tag{2.13}$$

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:

☞ sect. 21.2.1

$$x = (\mathbf{p}, \mathbf{q})\,, \qquad \mathbf{q} = (q_1, q_2, \ldots, q_D)\,, \qquad \mathbf{p} = (p_1, p_2, \ldots, p_D)\,. \tag{2.14}$$

In chapter 23 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

☞ chapter 23

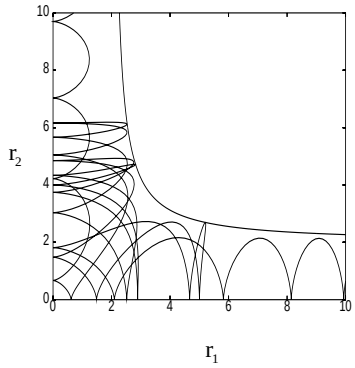


Figure 2.4: A typical colinear helium trajectory in the $r_1 - r_2$ plane; the trajectory enters here along the r_1 axis and then, like almost every other trajectory, after a few bounces escapes to infinity, in this case along the r_2 axis.

$$H = \frac{1}{2}p_1^2 + \frac{1}{2}p_2^2 - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_1 + r_2}\,. \tag{2.15}$$

The collinear helium has 2 degrees of freedom, thus a 4-dimensional phase space $\mathcal{M}$, which the energy conservation reduces to 3 dimensions. The dynamics can be visualized as a motion in the (r_1, r_2), $r_i \geq 0$ quadrant, fig. 2.4. It looks messy, and indeed it will turn out to be no less chaotic than a pinball bouncing between three disks.

fast track:
chapter 2.4, p. 44

2.3 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 phase space and changing the coordinates of phase space affects the coordinates of the tangent space as well.

We shall denote by h the *conjugation function* which maps the coordinates of the initial phase space manifold $\mathcal{M}$ into the reparametrized phase space manifold $\mathcal{M}'$, with a point $x \in \mathcal{M}$ related to a point $y \in \mathcal{M}'$ by $y = h(x)$. 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 $x = h^{-1}(y)$. As we interested in representing smooth flows, the reparametrized dynamics should support the same number of derivatives as the initial one. Ideally h is a (piece-wise) analytic function, in which case we refer to h as a *smooth conjugacy*.

The evolution rule $g^t(y_0)$ on the manifold $\mathcal{M}'$ can be computed from the evolution rule $f^t(x_0)$ on $\mathcal{M}$ and the coordinate change h. Take a point on $\mathcal{M}'$,

go back to $\mathcal{M}$, evolve, and then return to $\mathcal{M}'$:

$$y(t) = g^t(y_0) = h \circ f^t \circ h^{-1}(y_0)\,. \tag{2.16}$$

The vector field $v(x)$ locally tangent the flow f^t, found by differentiation (2.5), defines the flow $\dot{x} = v(x)$ in $\mathcal{M}$. The vector field $w(y)$ tangent to g^t which describes the flow $\dot{y} = w(y)$ in $\mathcal{M}'$ follows by differentiation and application of the chain rule:

$$\begin{aligned} w(y) &= \frac{\partial g^0}{\partial t}(0) = \frac{\partial}{\partial t} h \circ f^t \circ h^{-1}(y)\Big|_{y,t=0} \\ &= h'(h^{-1}(y))v(h^{-1}(y)) = h'(x)v(x)\,. \end{aligned} \tag{2.17}$$

2.6 on p. 53

The change of coordinates has to be a smooth one-to-one function, with h preserving the topology of the flow, or the manipulations we just carried out would not hold. Trajectories that are closed loops in $\mathcal{M}$ will remain closed loops in the new manifold $\mathcal{M}'$, and so on.

Imagine the phase space made out of a rubber sheet with the vector field drawn on it. A coordinate change corresponds to pulling and tugging on the rubber sheet. Globally h deforms the rubber sheet $\mathcal{M}$ into $\mathcal{M}'$ in a highly nonlinear manner, but locally it simply rescales and deforms the tangent field by $\partial_j h_j$, hence the simple transformation law (2.17) for the velocity fields. However, we do need to insist on (sufficient) smoothness of h in order to preclude violent and irreversible acts such as cutting, glueing, or self-intersections of the distorted rubber sheet. Time itself is but one possible parametrization of the points along a trajectory, and it can also be redefined, $s = s(t)$, with the attendent modification of (2.17).

What we *really* care about is pinning down an invariant notion 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 group 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.

2.3.1 Rectification of flows

A profitable way to exploit invariance 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 localy 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 x_q for which $v(x_q) = 0$. 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 coordinates takes the standard form

$$\begin{aligned} \dot{y}_1 &= 1 \\ \dot{y}_2 = \dot{y}_3 &= \cdots = \dot{y}_d = 0\,, \end{aligned} \tag{2.18}$$

with unit velocity flow along y_1, and no flow along any of the remaining directions.

2.3.2 Harmonic oscillator, rectified

As a simple example of global rectification of a flow consider the harmonic oscillator

$$\dot{q} = p\,, \qquad \dot{p} = -q\,. \tag{2.19}$$

The trajectories $x(t) = (p(t), q(t))$ just go around the origin, so a fair guess is that the system would have a simpler representation in polar coordinates $y = (r, \theta)$:

$$h^{-1} : \left\{ \begin{array}{rcl} q & = & h_1^{-1}(r,\theta) = r\cos\theta \\ p & = & h_2^{-1}(r,\theta) = r\sin\theta \end{array} \right. . \tag{2.20}$$

The Jacobian matrix of the transformation is

$$h' = \begin{bmatrix} \cos\theta & \sin\theta \\ -\dfrac{\sin\theta}{r} & -\dfrac{\cos\theta}{r} \end{bmatrix} \tag{2.21}$$

resulting in (2.17)

$$\dot{r} = 0\,, \qquad \dot{\theta} = -1\,. \tag{2.22}$$

In the new coordinates the radial coordinate r is constant, and the angular coordinate θ wraps around a cylinder with constant angular velocity. There is a subtle point in this change of coordinates: the domain of the map h^{-1} is not the the whole plane $\mathbb{R}^2$, but rather the whole 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 stationary point $x = (0, 0)$, or $r = 0$.

2.3.3 Colinear helium, regularized

Though very simple in form, the Hamiltonian (2.15) is not the most convenient for numerical investigations of the system. In the (r_1, r_2) coordinates the potential is singular for $r_i \to 0$ nucleus-electron collisions, with velocity diverging to ∞. These 2-body collisions can be regularized by a rescaling of the time and the coordinates $(r_1, r_2, p_1, p_2) \to (Q_1, Q_2, P_1, P_2)$, in a manner to be described in chapter 23. For the purpose at hand it is sufficient to state the result: In the rescaled coordinates the equations of motion are

$$\begin{aligned} \dot{P}_1 &= 2Q_1\left[2 - \frac{P_2^2}{8} - Q_2^2\left(1 + \frac{Q_2^2}{R^4}\right)\right]; \qquad \dot{Q}_1 = \frac{1}{4}P_1 Q_2^2 \\ \dot{P}_2 &= 2Q_2\left[2 - \frac{P_1^2}{8} - Q_1^2\left(1 + \frac{Q_1^2}{R^4}\right)\right]; \qquad \dot{Q}_2 = \frac{1}{4}P_2 Q_1^2 \,. \end{aligned} \tag{2.23}$$

where $R = (Q_1^2 + Q_2^2)^{1/2}$. These equations look harder to tackle than the harmonic oscillators that you are familiar with from other learned treatises, and indeed they are. But they are also a typical example of kinds of flows that one works with in practice, and the skill required in finding a good re-coordinatization $h(x)$.

in depth:
chapter 23, p. 529

2.4 Computing trajectories

You have not learned dynamics unless you know how to integrate numerically whatever dynamical equations you face. Stated tersely, 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:

$$x_i \to x_i + \delta\tau v_i(x)\,. \tag{2.24}$$

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-Kuta code in some mundane everyday language. While you absolutely need to master the requisite numerical methods, this in not the time or place to expand on them; how you learn them is your business.

2.8 on p. 54
2.9 on p. 54
2.10 on p. 54

And if you have developed some nice routines for solving problems in this text or can point another students to some, let us know.

2.11 on p. 55

fast track:
chapter 3, p. 57

2.5 Infinite-dimensional flows

Flows described by partial differential equations are considered infinite dimensional because if one writes them down as a set of ordinary differential equations (ODE) then one needs an infinity of the ordinary kind to represent the dynamics of one equation of the partial kind (PDE). Even though the phase space is infinite dimensional, for many systems of physical interest the global attractor is finite dimensional. We illustrate how this works with a concrete example, the Kuramoto-Sivashinsky system.

2.5.1 Partial differential equations

First, a few words about partial differential equations in general. Many of the partial differential equations of mathematical physics can be written in the quasi-linear form

$$\partial_t u = \mathbf{A}u + N(u)\,, \tag{2.25}$$

where u is a function (possibly a vector function) of the coordinate x and time t, $\mathbf{A}$ is a linear operator, usually containing the Laplacian and a few other derivatives of u, and $N(u)$ is the nonlinear part of the equation (terms like $u\partial_x u$ in (2.31) below).

Not all equations are stated in the form (2.25), but they can easily be so transformed, just as the ordinary differential equations can be rewritten as first-order systems. We will illustrate the method with a variant of the D'Alambert's wave equation describing a plucked string:

$$\partial_{tt} y = \left(c + \frac{1}{2} (\partial_x y)^2 \right) \partial_{xx} y \tag{2.26}$$

Were the term $\partial_x y$ small, this equation would be just the ordinary wave equation. To rewrite the equation in the first order form (2.25), we need a field $u = (y, w)$ that is two-dimensional,

$$\partial_t \begin{bmatrix} y \\ w \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ c\partial_{xx} & 0 \end{bmatrix} \begin{bmatrix} y \\ w \end{bmatrix} + \begin{bmatrix} 0 \\ \partial_{xx} y (\partial_x y)^2/2 \end{bmatrix} . \tag{2.27}$$

The [2×2] matrix is the linear operator $\mathbf{A}$ and the vector on the far right is the nonlinear function $N(u)$. Unlike ordinary functions, differentiations are part of the function. The nonlinear part can also be expressed as a function on the infinite set of numbers that represent the field, as we shall see in the Kuramoto-Sivashinsky example (2.31).

The usual technique for solving the linear part is to use Fourier methods. Just as in the ordinary differential equation case, one can integrate the linear part of

☞ chapter 4.2

$$\partial_t u = \mathbf{A} u \tag{2.28}$$

to obtain

$$u(x,t) = e^{t\mathbf{A}} u(x,0) \tag{2.29}$$

If u is expressed as Fourier series $\sum_k a_k \exp(ikx)$, as we will do for the Kuramoto-Shivashinsky system, then we can determine the action of $e^{t\mathbf{A}}$ on $u(x,0)$. This can be done because differentiations in $\mathbf{A}$ act rather simply on the exponentials. For example,

$$e^{t\partial_x} u(x,0) = e^{t\partial_x} \sum_k a_k e^{ikx} = \sum_k a_k \frac{(it)^k}{k!} e^{ikx} . \tag{2.30}$$

Depending on the behavior of the linear part, one distinguishes three classes of partial differential equations: *diffusion, wave, and potential.* The classification relies on the solution by a Fourier series, as in (2.29). In mathematical literature

these equations are also called *parabolic, hyperbolic* and *elliptic*. If the nonlinear part $N(u)$ is as big as the linear part, the classification is not a good indication of behavior, and one can encounter features of one class of equations while studying the others.

In *diffusion-type* equations the modes of high frequency tend to become smooth, and all initial conditions tend to an attractor, called the *inertial manifold*. The Kuramoto-Sivashinsky system studied below is of this type. The solution being attracted to the inertial manifold does not mean that the amplitudes of all but a finite number of modes go to zero (alas were we so lucky), but that there is a finite set of modes that could be used to describe any solution of the inertial manifold. The only catch is that there is no simple way to discover what these inertial manifold modes might be.

In *wave-like* equations the high frequency modes do not die out and the solutions tend to be distributions. The equations can be solved by variations on the WKB idea: the wave-like equations can be approximated by the trajectories of the wave fronts.

chapter 21

2.12 on p. 56

Elliptic equations have no time dependence and do not represent dynamical systems.

2.5.2 Fluttering flame front

> Romeo: 'Misshapen chaos of well seeming forms!'
> W. Shakespeare, *Romeo and Julliet*, act I, scene I

The Kuramoto-Sivashinsky equation, arising in description of the flame front flutter of gas burning in a cylindrically symmetric burner on your kitchen stove and many other problems of greater import, is one of the simplest partial differential equations that exhibit chaos. It is a dynamical system extended in one spatial dimension, defined by

$$u_t = (u^2)_x - u_{xx} - \nu u_{xxxx} \,. \tag{2.31}$$

In this equation $t \geq 0$ is the time and $x \in [0, 2\pi]$ is the space coordinate. The subscripts x and t denote the partial derivatives with respect to x and t; $u_t = du/dt$, u_{xxxx} stands for 4th spatial derivative of the "height of the flame front" (or perhaps "velocity of the flame front") $u = u(x,t)$ at position x and time t. ν is a "viscosity" parameter; its role is to suppress solutions with fast spatial variations. The term $(u^2)_x$ makes this a *nonlinear system*. It is the simplest conceivable PDE nonlinearity, playing the role in applied mathematics analogous to the role that the x^2 nonlinearity (3.11) plays in the dynamics of iterated

mappings. Time evolution of a solution of the Kuramoto-Sivashinsky system is illustrated by fig. 2.5. How are such solutions computed? The salient feature of such partial differential equations is that for any finite value of the phase-space contraction parameter ν a theorem says that the asymptotic dynamics is describable by a *finite* set of "inertial manifold" ordinary differential equations.

The "flame front" $u(x,t) = u(x+2\pi,t)$ is periodic on the $x \in [0, 2\pi]$ interval, so a reasonable strategy (but by no means the only one) is to expand it in a discrete spatial Fourier series:

$$u(x,t) = \sum_{k=-\infty}^{+\infty} b_k(t) e^{ikx} \,. \tag{2.32}$$

Since $u(x,t)$ is real, $b_k = b^*_{-k}$. Substituting (2.32) into (2.31) yields the infinite ladder of evolution equations for the Fourier coefficients b_k:

$$\dot{b}_k = (k^2 - \nu k^4) b_k + ik \sum_{m=-\infty}^{\infty} b_m b_{k-m} \,. \tag{2.33}$$

As $\dot{b}_0 = 0$, the solution integrated over space is constant in time. In what follows we shall consider only the cases where this average is zero, $b_0 = \int dx\, u(x,t) = 0$.

Coefficients b_k are in general complex functions of time t. We can simplify the system (2.33) further by considering the case of b_k pure imaginary, $b_k = ia_k$, where a_k are real, with the evolution equations

$$\dot{a}_k = (k^2 - \nu k^4) a_k - k \sum_{m=-\infty}^{\infty} a_m a_{k-m} \,. \tag{2.34}$$

This picks out the subspace of odd solutions $u(x,t) = -u(-x,t)$, so $a_{-k} = -a_k$.

That is the infinite set of ordinary differential equations promised at the beginning of the section.

The trivial solution $u(x,t) = 0$ is an equilibrium point of (2.31), but that is basically all we know as far as analytical solutions are concerned. You can integrate numerically the Fourier modes (2.34), truncating the ladder of equations to a finite number of modes N, that is, set $a_k = 0$ for $k > N$. In applied mathematics literature this is called a *Galerkin truncation*. For parameter values explored below, $N \leq 16$ truncations were deemed sufficiently accurate. If your integration routine takes days and lots of memory, you should probably

2.8 on p. 54

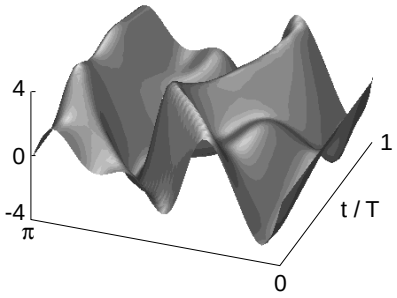


Figure 2.5: Spatiotemporally periodic solution $u_0(x,t)$. We have divided x by π and plotted only the $x > 0$ part, since we work in the subspace of the odd solutions, $u(x,t) = -u(-x,t)$. $N = 16$ Fourier modes truncation with $\nu = 0.029910$. (From ref. [6])

start from scratch and write a few lines of your own Runge-Kuta code.

Once the trajectory is computed in the Fourier space, we can recover and plot the corresponding spatiotemporal pattern $u(x,t)$ over the configuration space using (2.32), as in fig. 2.5.

2.5.3 Fourier modes truncations

The growth of the unstable long wavelengths (low $|k|$) excites the short wavelengths through the nonlinear term in (2.34). The excitations thus transferred are dissipated by the strongly damped short wavelengths, and a sort of "chaotic equilibrium" can emerge. The very short wavelengths $|k| \gg 1/\sqrt{\nu}$ will remain small for all times, but the intermediate wavelengths of order $|k| \sim 1/\sqrt{\nu}$ will play an important role in maintaining the dynamical equilibrium. Hence, while one may truncate the high modes in the expansion (2.34), care has to be exercised to ensure that no modes essential to the dynamics are chopped away. In practice one does this by repeating the same calculation at different truncation cutoffs N, and making sure that inclusion of additional modes has no effect within the accuracy desired. For figures given here, the numerical calculations were performed taking $N = 16$ and the damping parameter value $\nu = 0.029910$, for which the system is chaotic (as far as we can determine that numerically).

The problem with such high dimensional truncations of the infinite tower of equations (2.34) is that the dynamics is difficult to visualize. The best we can do without much programming (thinking is extra price) is to examine the trajectory's projections onto any three axes a_i, a_j, a_k, as in fig. 2.6.

☞ sect. 3.1.2

We can now start to understand the remark on page 37 that for infinite dimensional systems time reversability is not an option: evolution forward in time strongly damps the higher Fourier modes. But if we reverse the time, the infinity of high modes that contract strongly forward in time now explodes, rendering evolution backward in time meaningless.

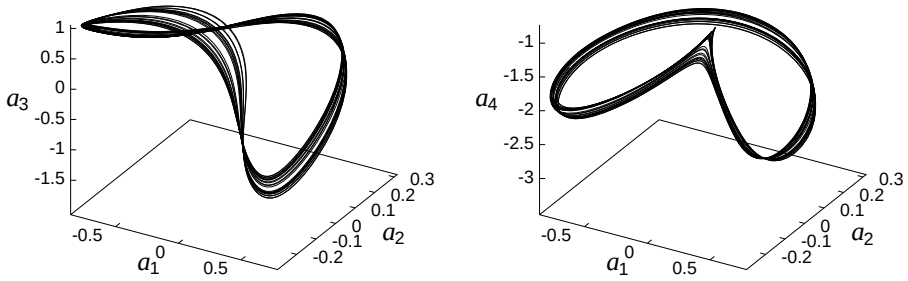


Figure 2.6: Projections of a typical 16-dimensional trajectory onto different 3-dimensional subspaces, coordinates (a) $\{a_1, a_2, a_3\}$, (b) $\{a_1, a_2, a_4\}$. $N = 16$ Fourier modes truncation with $\nu = 0.029910$. (From ref. [6].)

Commentary

Remark **2.1** Rössler, Kuramoto-Shivashinsky, and PDE systems. Rössler system was introduced in ref. [2], as a simplified set of equations describing time evolution of concentrations of chemical reagents. The Duffing system (2.7) arises in study of electronic circuits.The theorem on finite dimenionality of inertial manifolds of phase-space contracting PDE flows is proven in ref. [3]. The Kuramoto-Sivashinsky equation was introduced in ref. [4, 5]; sect. 2.5 is based on V. Putkaradze's term project paper (see `www.nbi.dk/-ChaosBook/`extras/), and Christiansen *et al.* [6]. How good description of a flame front this equation is need not concern us here; suffice it to say that such model amplitude equations for interfacial instabilities arise in a variety of contexts - see e.g. ref. [7] - and this one is perhaps the simplest physically interesting spatially extended nonlinear system.

Résumé

A *dynamical system* – a flow, a return map constructed from a Poincaré section of the flow, or an iterated map – is defined by specifying a pair $(\mathcal{M}, f)$, where $\mathcal{M}$ is a space and $f : \mathcal{M} \to \mathcal{M}$. The key concepts in exploration of the long time dynamics are the notions of *recurrence* and of the *non–wandering set* of f, the union of all the non-wandering points of $\mathcal{M}$. In more visual terms, chaotic dynamics with a low dimensional attractor can be thought of as a succession of nearly periodic but unstable motions.

Similarly, 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 alphabet of admissible patterns, and the long term dynamics can be thought of as a walk through the space of such patterns.

References

[2.1] E.N. Lorenz, *J. Atmospheric Phys.* **20**, 130 (1963).

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

[2.3] See e.g. Foias C, Nicolaenko B, Sell G R and Témam R Kuramoto-Sivashinsky equation *J. Math. Pures et Appl.* **67** 197, (1988).

[2.4] Kuramoto Y and Tsuzuki T Persistent propagation of concentration waves in dissipative media far from thermal equilibrium *Progr. Theor. Physics* **55** 365, (1976).

[2.5] Sivashinsky G I Nonlinear analysis of hydrodynamical instability in laminar flames - I. Derivation of basic equations *Acta Astr.* **4** 1177, (1977).

[2.6] F. Christiansen, P. Cvitanović and V. Putkaradze, *"Spatiotemporal chaos in terms of unstable recurrent patterns"*, *Nonlinearity* **10**, 55 (1997), `chao-dyn/9606016`

[2.7] Kevrekidis I G, Nicolaenko B and Scovel J C Back in the saddle again: a computer assisted study of the Kuramoto-Sivashinsky equation *SIAM J. Applied Math.* **50** 760, (1990).

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

Exercises

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

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

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

2.2 Evolution as a group. The trajectory evolution f^t is a one-parameter group where

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

Show that it is a commutative group.

In this case, the commutative character of the group of evolution functions comes from the commutative character of the time parameter under addition. Can you see any other 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 are a simple dynamical systems where the velocity field is given by the gradient of an auxiliary function ϕ

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

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

(a) Show that the velocity of the particle is in the direction of most rapid decrease of the function ϕ.

(b) Show that all extrema of ϕ are fixed points of the flow.

(c) Show that it takes an infinite amount of time for the system to reach an equilibrium point.

(d) Show that there are no periodic orbits in gradient systems.

2.6 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

$$\frac{d}{dt}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 5y_1 + 2y_2 \\ y_1 + 4y_2 \end{bmatrix}\,. \tag{2.35}$$

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

2.7 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 αz, that is,

$$f(z) = h^{-1}(\alpha h(z))\,.$$

There are conditions on the derivative of f at the origin to assure that the conjugation is always possible. Can you formulate these conditions by examining the series?

(difficulty: medium)

2.8 Runge-Kutta integration. Implement the fourth-order Runge-Kutta integration formula (see, for example, ref. [8]) for $\dot{x} = v(x)$:

$$\begin{aligned} x_{n+1} &= x_n + \frac{k_1}{6} + \frac{k_2}{3} + \frac{k_3}{3} + \frac{k_4}{6} + O(\delta\tau^5) \\ k_1 &= \delta\tau v(x_n)\,, \qquad k_2 = \delta\tau v(x_n + k_1/2) \\ k_3 &= \delta\tau v(x_n + k_2/2)\,, \quad k_4 = \delta\tau v(x_n + k_3) \end{aligned} \tag{2.36}$$

or some other numerical integration routine.

2.9 Rössler system. Use the result of exercise 2.8 or some other integration routine to integrate numerically the Rössler system (2.12). Does the result look like a "strange attractor"?

2.10 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 $\dot{x} = x^2\,, \qquad x(0) = 1\,.$

(a) For what times do solutions of

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

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

(b) Let's test the integrator you wrote in exercise 2.8. The equation $\ddot{x} = -x$ with initial conditions $x(0) = 2$ and $\dot{x} = 0$ has as solution $x(t) = e^{-t}(1+e^{2\,t})$. 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 noticeble (the exponential growth factor is ≈ 2.4).

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

2.11 Classical collinear helium dynamics. In order to apply the 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 (2.15)

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

The nuclear charge for helium is $Z = 2$. The 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 the (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. 23.1. In the transformed coordinates (x_1, x_2, p_1, p_2) the Hamiltonian equations of motion take the form (2.23).

(a) Integrate the equations of motion by the fourth order Runge-Kutta computer routine of exercise 2.8 (or whatever integration routine you like). A convenient way to visualize the 3-d phase space orbit is by projecting it onto the 2-dimensional $(r_1(t), r_2(t))$ plane.

(Gregor Tanner, Per Rosenqvist)

2.12 Infinite dimensional dynamical systems are not smooth. Many of the operations we consider natural for finite dimensional systems do not have not smooth behavior in infinite dimensional vector spaces. Consider, as an example, a concentration ϕ diffusing on $\mathbb{R}$ according to the diffusion equation

$$\partial_t \phi = \frac{1}{2}\nabla^2 \phi\,.$$

(a) Interpret the partial differential equation as an infinite dimensional dynamical system. That is, write it as $\dot{x} = F(x)$ and find the velocity field.

(b) Show by examining the norm

$$\|\phi\|^2 = \int_{\mathbb{R}} dx\, \phi^2(x)$$

that the vector field F is not continuous.

(c) Try the norm

$$\|\phi\| = \sup_{x\in\mathbb{R}} |\phi(x)| \,.$$

Is F continuous?

(d) Argue that the semi-flow nature of the problem is not the cause of our difficulties.

(e) Do you see a way of generalizing these results?

Chapter 3

Maps

(R. Mainieri and P. Cvitanović)

The time parameter in the definition of a dynamical system, sect. 2.1, can be either continuous or discrete. Discrete time dynamical systems arise naturally from flows; one can observe the flow at fixed time intervals (the strobe method), 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 having one of the coordinates become zero, or as complicated as having the flow cut through a curved hypersurface. There are also settings where discrete time is altogether natural, for example a particle moving through a billiard, sect. 3.4, suffers a sequence of instantaneous kicks, and executes a simple motion between successive kicks.

fast track:
chapter 5, p. 97

3.1 Poincaré sections

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

$$x_{n+1} = P(x_n)\,, \qquad x_m \in \mathcal{P}\,. \tag{3.1}$$

The choice of the section hypersurface $\mathcal{P}$ is altogether arbitrary. However, with a sufficiently clever choice of a Poincaré section or a set of sections, any orbit

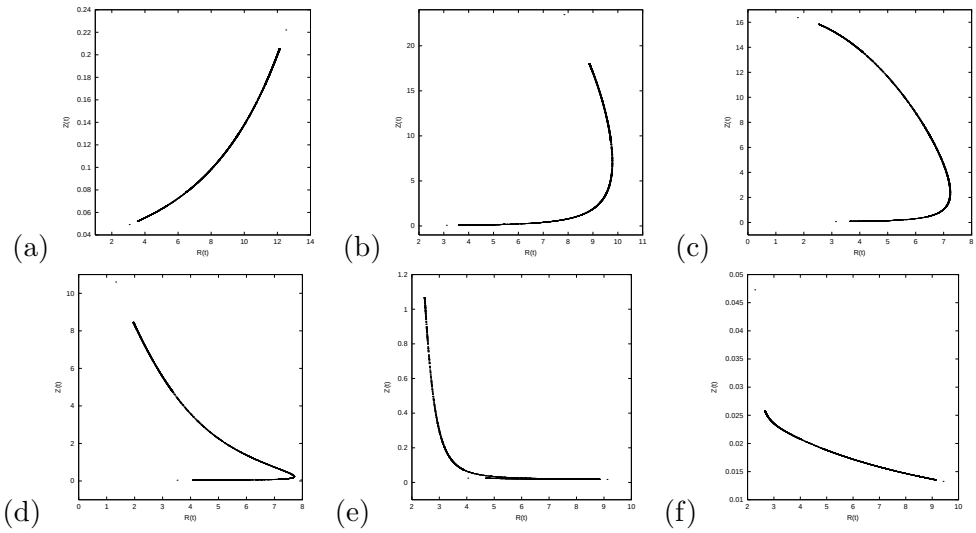


Figure 3.1: (b) Poincaré sections of the Rössler flow at $t = 20000$ taken with a plane through z-axis, at angles (a) 135^o, (b) 90^o, (c) 45^o, (d) 0^o, (e) 315^o, (f) 270^o in the x-y-plane. This sequence of Poincaré sections illustrates the "stretch & fold" action of the Rössler flow. To orient yourself, compare with fig. 2.3, and note the different z-axis scales. The segment starts out close to the x-y plane, and after the folding (b) $\to$ (c) $\to$ (d) the folded segment is returned close to the x-y plane strongly compressed. (G. Simon)

of interest intersects a section. Depending on the application, one might need to supplement the return map with the time of *first return function* $\tau(x_n)$ - sometimes refered to as *ceiling function* - which gives the time of flight to the next section for a trajectory starting at x_n, with the accumulated flight time given by

$$t_{n+1} = t_n + \tau(x_n)\,, \qquad t_0 = 0\,, \qquad x_n \in \mathcal{P}\,. \tag{3.2}$$

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

☞ chapter 6

An example may help visualize this. Consider the simple pendulum. Its phase space is 2-dimensional: momentum on the vertical axis and position on the horizontal axis. We can then choose the Poincaré section to be the positive horizontal axis. Now imagine what happens as a point traces a trajectory through this phase space. 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 dissipation. Now every trajectory is an inwards spiral, and the trajectory will intersect the Poincaré section at a series of points that get closer and closer to the origin.

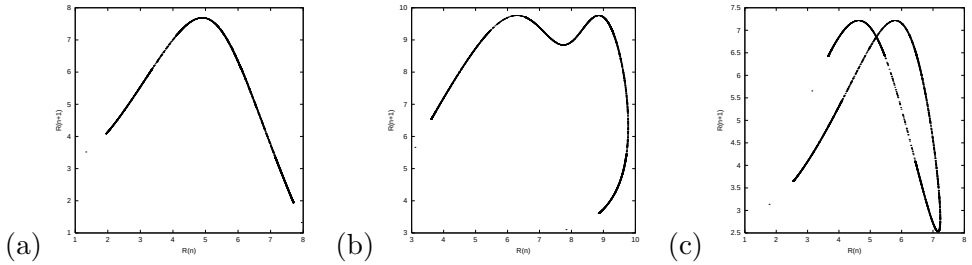


Figure 3.2: Return maps for the $R_n \to R_{n+1}$ radial distance constructed from different Poincaré sections for the Rössler flow, at angles (a) 0^o, (b) 90^o, (c) 45^o around the z-axis, see fig. 3.1. The case (a) is an example of a nice 1-to-1 return map. However, (b) and (c) appear multimodal and non-invertible. These are artifacts of projections of a 2-dimensional return map $(R_n, z_n) \to (R_{n+1}, z_{n+1})$ onto a 1-dimensional subspace $R_n \to R_{n+1}$. (G. Simon)

3.1.1 A Poincaré map with a strange attractor

Appreciation of the utility of visualization of dynamics by means of Poincaré sections is gained through experience. Consider a 3-dimensional visualization of the Rössler flow (2.12), such as fig. 2.3. The trajectories seem to wrap around the origin, so a good choice for a Poincaré section may be a plane containing the z axis. Fig. 3.1 illustrates what the Poincaré sections containing the z axis and oriented at different angles with respect to the x axis look like. Once the section is fixed, we can construct a return map (3.1), as in fig. 3.2. A Poincaré section gives us a much more informative snapshot of the flow than the full flow portrait; for example, we see in the Poincaré section that even though the return map is 2-$d \to$ 2-d, for the Rössler system the flow contraction happens to be so strong that for all practical purposes it renders the return map 1-dimensional.

fast track:
sect. 3.3, p. 62

3.1.2 Fluttering flame front

One very human problem with dynamics such as the high-dimensional truncations of the infinite tower of the Kuramoto-Sivashinsky modes (2.34) is that the dynamics is difficult to visualize.

The question is how to look at such flow? One of the first steps in analysis of such flows is to restrict the dynamics to a Poincaré section. We fix (arbitrarily) the Poincaré section to be the hyperplane $a_1 = 0$, and integrate (2.34) with the initial conditions $a_1 = 0$, and arbitrary values of the coordinates $a_2, \ldots, a_N$, where N is the truncation order. When a_1 becomes 0 the next time and the flow

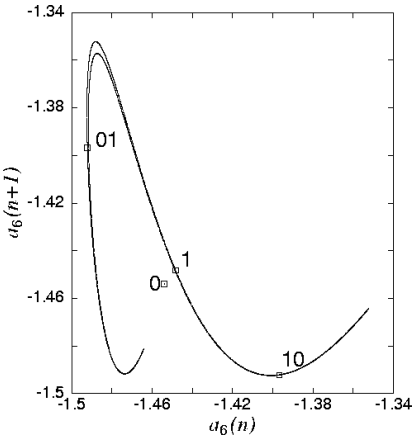


Figure 3.3: The attractor of the Kuramoto-Sivashinsky system (2.34), plotted as the a_6 component of the $a_1 = 0$ Poincaré section return map. Here 10,000 Poincaré section returns of a typical trajectory are plotted. Also indicated are the periodic points $\overline{0}$, $\overline{1}$ and $\overline{01}$. $N = 16$ Fourier modes truncation with $\nu = 0.029910$. (From ref. [6].)

crosses the hyperplane $a_1 = 0$ in the same direction as initially, the coordinates $a_2, \ldots, a_N$ are mapped into $(a'_2, \ldots a'_N) = P(a_2, \ldots, a_N)$, where P is the Poincaré mapping of the $(N-1)$-dimensional $a_1 = 0$ hyperplane into itself. Fig. 3.3 is an example of a result that one gets. We have to pick - arbitrarily - a subspace such as a'_6 vs. a_6 to visualize the dynamics. While the topology of the attractor is still obscure, one thing is clear - the attractor is finite and thin, barely thicker than a line.

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.

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

$$\frac{dx_i}{dt} = v_i(x,t)\,, \tag{3.3}$$

where the flow velocity v is a vector function of the position in phase space x and the time t. In general v cannot be integrated analytically and 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 surface. The surface will be specified implicitly through a function $g(x)$ that is zero whenever a point x is on the Poincaré section. The simplest choice of such section is a plane 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 on this plane if it satisfies the condition

3.2 on p. 69

$$g(x) = (x - r_0) \cdot a = 0\,. \tag{3.4}$$

If we use a tiny step size in our numerical integrator, we can observe the value of g as we integrate; its sign will change as the trajectory crosses the surface. The problem with this method is that we have to use a very small integration time step. In order to actually land on the Poincaré section one might try to interpolate the intersection point from the two trajectory points on either side of the surface. However, there is a better way.

Let t_a be the time just before g 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. Suppose that x_1 is not tangent to the Poincaré section. Using

$$\frac{dx_k}{dx_1}\frac{dx_1}{dt} = \frac{dx_k}{dx_1}v_1(x,t) = v_k(x,t) \tag{3.5}$$

we can rewrite the equations of motion (3.3) as

$$\begin{aligned} \frac{dt}{dx_1} &= \frac{1}{v_1} \\ &\vdots \\ \frac{dx_k}{dx_1} &= \frac{v_k}{v_1}\,. \end{aligned} \tag{3.6}$$

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 surface, which can be found from the surface intersection condition (3.4). x_1 need not be perpendicular to the Poincaré section; any x_i can be picked as the integration variable, as long as the x_i axis is not parallel to the Poincaré section.

The functional form of $P(x)$ can be obtained by tabulating the results of integration of the flow from x to the first Poincaré section return for many $x \in \mathcal{P}$, and interpolating. It might pay to find a good approximation to $P(x)$, and then get rid of numerical integration altogether by replacing $f^t(x)$ by iteration of the Poincaré return map $P(x)$. Polynomial approximations

$$P_k(x) = a_k + \sum_{j=1}^{d} b_{kj}x_j + \sum_{i,j=1}^{d} c_{kij}x_i x_j + \dots\,, \qquad x \in \mathbb{R}^n \tag{3.7}$$

to Poincaré return maps

$$\begin{pmatrix} x_{1,n+1} \\ x_{2,n+1} \\ \cdots \\ x_{d,n+1} \end{pmatrix} = \begin{pmatrix} P_1(x_n) \\ P_2(x_n) \\ \cdots \\ P_d(x_n) \end{pmatrix} , \qquad n\text{th Poincaré section return} ,$$

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

3.3 Hénon map

The example of a nonlinear 2-dimensional map most frequently employed in testing various hunches about chaotic dynamics, the "E. Coli" of nonlinear dynamics, is the *Hénon map*

$$\begin{aligned} x_{n+1} &= 1 - ax_n^2 + by_n \\ y_{n+1} &= x_n , \end{aligned} \tag{3.8}$$

sometimes written equivalently as the 2-step recurrence relation

$$x_{n+1} = 1 - ax_n^2 + bx_{n-1} . \tag{3.9}$$

Parenthetically, an n-step recurrence relation is the discrete time analogue of nth order differential equation, and it can always be replaced by a set of 1-step recurrence relations. Another example frequently employed is the *Lozi map*, a linear, "tent map" version of the Hénon map given by

$$\begin{aligned} x_{n+1} &= 1 - a|x_n| + by_n \\ y_{n+1} &= x_n . \end{aligned} \tag{3.10}$$

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 whole class of maps of the Hénon type, so called once-folding maps.

The Hénon map is the simplest map that captures the "stretch & fold" dynamics of return maps such as the Rössler's, fig. 3.2(a). It can be obtained by a truncation of a polynomial approximation (3.7) to a Poincaré return map to second order.

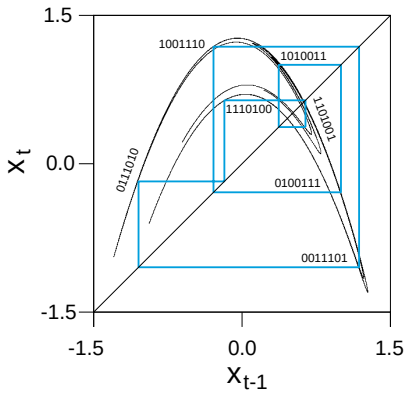


Figure 3.4: The strange attractor (unstable manifold) and a period 7 cycle of the Hénon map (3.8) with $a = 1.4$, $b = 0.3$. The periodic points in the cycle are connected to guide the eye; for a numerical determination of such cycles, consult sect. 12.4.1. (K.T. Hansen)

The Hénon map dynamics is conveniently plotted in the (x_n, x_{n+1}) plane; an example is given in fig. 3.4. A quick sketch of asymptotics of such mapping is obtained by picking an arbitrary starting point and iterating (3.8) on a computer. For an arbitrary initial point this process might converge to a stable limit cycle, to a strange attractor, to a false attractor (due to the roundoff errors), or diverge. In other words, straight iteration is essentially uncontrollable, and we will need to resort to more thoughtful explorations. As we shall explain in due course below, strategies for systematic exploration rely on stable/unstable manifolds, periodic points, saddle-stradle methods and so on.

3.4 on p. 70

3.5 on p. 70

The Hénon map stretches out and folds once a region of the (x, y) plane centered around the origin. Parameter a controls the amount of stretching, while parameter b controls the thickness of the folded image through the "1-step memory" term bx_{n-1} in (3.9), see fig. 3.4. For small b the Hénon map reduces to the 1-dimensional quadratic map

$$x_{n+1} = 1 - ax_n^2 \,. \tag{3.11}$$

3.6 on p. 70

By setting $b = 0$ we lose determinism, as (3.11) inverted has two preimages $\{x_{n+1}^+, x_{n+1}^-\}$ for most x_n. Still, the approximation is very instructive. As we shall see in sect. 10.5, understanding of 1-dimensional dynamics is indeed the essential prerequisite to unravelling 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.

fast track:
chapter 4, p. 73

We note here a few simple symmetries of the Hénon maps for future reference. For $b \neq 0$ the Hénon map is reversible: the backward iteration of (3.9) is given

by

$$x_{n-1} = -\frac{1}{b}(1 - ax_n^2 - x_{n+1})\,. \tag{3.12}$$

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 (see (15.1) below),

$$x_{n-1} = 1 - ax_n^2 - x_{n+1}\,, \tag{3.13}$$

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

$$x_{n-1} = 1 - ax_n^2 + x_{n+1}\,, \tag{3.14}$$

and the non–wandering set is symmetric across the $x_{n+1} = -x_n$ diagonal.

3.4 Billiards

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 $\mathbb{R}^D/Q$. In what follows we shall more often than not restrict our attention to $D = 2$ planar billiards. A point particle ("pinball") of mass m and momentum $p_i = mv_i$ moves freely within the billiard, along a straight line, until it encounters the boundary. There it reflects specularly, with instantaneous change in the momentum component orthogonal to the boundary,

$$\overrightarrow{p}' = \overrightarrow{p} - 2(\overrightarrow{p} \cdot \hat{n})\hat{n}\,, \tag{3.15}$$

where $\hat{n}$ is a unit vector normal to the boundary ∂Q at the collision point. The angle of incidence equals to the angle of reflection. A billiard is a Hamiltonian system with a $2D$-dimensional phase space $x = (p, q)$ and potential $V(q) = 0$ for $q \in Q$, and $V(q) = \infty$ for $q \in \partial Q$. Without loss of generality we will set $m = |v| = 1$ throughout.

If we know what happens at two successive collisions we can determine quite easily what happens in between, as the position of a point of reflection together

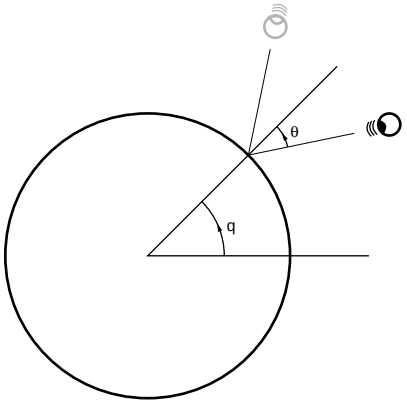


Figure 3.5: Angles defining a unique billiard trajectory. The coordinate q is given by an angle in $[0, 2\pi]$, and the momentum is given by specifying its component $\sin\theta$ tangential to the disk. For convenience, the pinball momentum is customarily set equal to one.

with the outgoing trajectory angle uniquely specifies the trajectory. In sect. 1.3.4 we used this observation to reduce the pinball flow to a map by the Poincaré section method, and associate an iterated mapping to the three-disk flow, a mapping that takes us from one collision to the next.

A billiard flow has a natural Poincaré section defined by marking q_i, the arc length position of the ith bounce measured along the billiard wall, and $p_i = \sin\phi_i$, the momentum component parallel to the wall, where ϕ_i is the angle between the outgoing trajectory and the normal to the wall. We measure the arc length q anti-clockwise relative to the interior of a scattering disk, see fig. 1.5(a). The dynamics is then conveniently described as a map $P : (q_n, p_n) \mapsto (q_{n+1}, p_{n+1})$ from the nth collision to the $(n+1)$th collision. Coordinates $x_n = (q_n, p_n)$ are the natural choice (rather than, let's say, (q_i, ϕ_i)), because they are phase-space volume preserving, and easy to extract from the pinball trajectory.

4.7 on p. 96

Let t_k be the instant of kth collision. Then the position of the pinball $\in Q$ at time $t_k + \tau \leq t_{k+1}$ is given by $2D-2$ Poincaré section coordinates $(q_k, p_k) \in \mathcal{P}$ together with τ, the distance reached by the pinball along the kth section of its trajectory. In $D = 2$, the Poincaré section is a cylinder where the parallel momentum p ranges for -1 to 1, and the q coordinate is cyclic along each connected component of ∂Q.

sect. 4.5

3.4.1 3-disk game of pinball

For example, for the 3-disk game of pinball of fig. 1.3 and fig. 1.5 we have two types of collisions:

3.7 on p. 71

$$P_0 : \begin{cases} \varphi' = -\varphi + 2\arcsin p \\ p' = -p + \frac{a}{R}\sin\varphi' \end{cases} \qquad \text{back-reflection} \tag{3.16}$$

$$P_1 : \begin{cases} \varphi' = \varphi - 2\arcsin p + 2\pi/3 \\ p' = p - \frac{a}{R}\sin\varphi' \end{cases} \qquad \text{reflect to 3rd disk}\,. \tag{3.17}$$

Actually, as in this case 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.

3.8 on p. 71

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 R by the local curvature of the wall at the point of collision.

Commentary

Remark **3.1** Hénon, Lozi maps. The Hénon map *per se* is of no special significance - its importance 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]. Y. Pomeau's studies of the Lorenz attractor on an analog computer, and his insights into its stretching and folding dynamics led Hénon [1] to the Hénon mapping in 1976. Hénon's and Lorenz's original papers can be found in reprint collections refs. [2, 3]. They are a pleasure to read, and are still the best introduction to the physics background motivating such models. Detailed description of the Hénon map dynamics was given by Mira and coworkers [4], as well as very many other authors.

The Lozi map [5] is particularly convenient in investigating the symbolic dynamics of 2-d mappings. Both the Lorenz and the Lozi system are uniformly smooth maps with singularities. For the Lozi maps the continuity of measure was proven by M. Misiurewicz [6], and the existence of the SRB measure was established by L.-S. Young.

Remark **3.2** 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 [7]. 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 cycle expansions curvature corrections that are central to this treatise.

chapter 13

That the 3-disk game of pinball is a quintessential example of deterministic chaos appears to have been first noted by B. Eckhardt [7]. The

model was studied in depth classically, semiclassically and quantum mechanically by P. Gaspard and S.A. Rice [8], and used by P. Cvitanović and B. Eckhardt [9] 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 [10], construct semiclassical evolution operators and entire spectral determinants by P. Cvitanović and G. Vattay [11], and incorporate the diffraction effects into the periodic orbit theory by G. Vattay, A. Wirzba and P.E. Rosenqvist [12]. The full quantum mechanics and semiclassics of scattering systems is developed here in the 3-disk scattering context in chapter **??**. Gaspard's monograph [4], which we warmly recommend, utilizies the 3-disk system in much more depth than will be attained here. For further links check `www.nbi.dk/ChaosBook/`.

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.

References

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

[3.2] *Universality in Chaos, 2. edition*, P. Cvitanović, ed., (Adam Hilger, Bristol 1989).

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

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

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

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

[3.7] B. Eckhardt, *Fractal properties of scattering singularities*, *J. Phys.* **A 20**, 5971 (1987).

[3.8] P. Gaspard and S.A. Rice, *J. Chem. Phys.* **90**, 2225 (1989); **90**, 2242 (1989); **90**, 2255 (1989).

[3.9] P. Cvitanović and B. Eckhardt, "Periodic-orbit quantization of chaotic system", *Phys. Rev. Lett.* **63**, 823 (1989).

[3.10] P. Gaspard and D. Alonso Ramirez, *Phys. Rev.* **A 45**, 8383 (1992).

[3.11] P. Cvitanović and G. Vattay, *Phys. Rev. Lett.* **71**, 4138 (1993).

[3.12] G. Vattay, A. Wirzba and P.E. Rosenqvist, *Periodic Orbit Theory of Diffraction*, *Phys. Rev. Lett.* **73**, 2304 (1994).

[3.13] C. Simo, in D. Baenest and C. Froeschlé, Les Méthodes Modernes de la Mécanique Céleste (Goutelas 1989), p. 285.

Exercises

3.1 Rössler system (continuation of exercise 2.9) Construct a Poincaré section for this flow. How good an approximation would a replacement of the return map for this section by a 1-dimensional map be?

3.2 Arbitrary Poincaré sections. We will generalize the construction of Poincaré section so that it can have any shape, as specified by the equation $g(x) = 0$.

(a) Start out 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

$$\frac{dx_k}{ds} = \kappa f_k \,, \tag{3.18}$$

with $dt/ds = \kappa$, and choosing κ to be 1 or $1/f_1$. This allows one to switch between t and x_1 as the integration "time."

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

$$x_{n+1} = g(x) \,. \tag{3.19}$$

How can this be used to find the Poincaré section?

3.3 Classical collinear helium dynamics. (continuation of exercise 2.11)

(a) 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 (2.15)). Compare your results with fig. 23.3(b).

(Gregor Tanner, Per Rosenqvist)

3.4 Hénon map fixed points. Show that the two fixed points (x_0, x_0), (x_1, x_1) of the Hénon map (3.8) are given by

$$\begin{aligned} x_0 &= \frac{-(1-b) - \sqrt{(1-b)^2 + 4a}}{2a}\,, \\ x_1 &= \frac{-(1-b) + \sqrt{(1-b)^2 + 4a}}{2a}\,. \end{aligned} \tag{3.20}$$

3.5 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" attracttor vanishes like a 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?

3.6 Fixed points of maps. A continuous function F is a contraction of the unit interval if it maps the interval inside itself.

(a) Use the continuity of F to show that a one-dimensional contraction F of the interval $[0, 1]$ has at least one fixed point.

(b) In a uniform (hyperbolic) contraction the slope of F is always smaller than one, $|F'| < 1$. Is the composition of uniform contractions a contraction? Is it uniform?

3.7 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 fig. 1.7 or fig. 10.3 for $R{:}a = 2.5$ and/or 6.

3.8 Trapped orbits. Shoot 100,000 trajectories from one of the disks, and trace out the strips of fig. 1.7 for various $R{:}a$ by color coding the initial points in the Poincaré section by the number of bounces preceeding 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.

Chapter 4

Local stability

(R. Mainieri and P. Cvitanović)

Topological features of a dynamical system – singularities, periodic orbits, and the overall topological interrelations between trajectories – are invariant under a general continuous change of coordinates. More surprisingly, there exist quantities that depend on the notion of metric distance between points, but nevertheless do not change value under a change of coordinates. Local quantities such as stability eigenvalues of equilibria and periodic orbits and global quantities such as the Lyapunov exponents, metric entropy, and fractal dimensions are examples of such coordinate choice independent properties of dynamical systems.

We now turn to our first class of such invariants, linear stability of flows and maps. This will give us metric information about local dynamics. Extending the local stability eigendirections into stable and unstable manifolds will yield important global information, a topological foliation of the phase space.

4.1 Flows transport neighborhoods

As a swarm of representative points moves along, it carries along and distorts neighborhoods, as sketched in fig. 2.1(b). 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 $x(t) = f^t(x_0)$. The system of linear *equations of variations* for the displacement of the infinitesimally close neighbor $x_i(x_0, t) + \delta x_i(x_0, t)$ follows from the flow equations (2.6) by Taylor expanding to

linear order

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

Taken together, the set of equations

$$\dot{x}_i = v_i(x)\,, \quad \dot{\delta x}_i = \mathbf{A}_{ij}(x)\delta x_j \tag{4.2}$$

governs the dynamics in the extended $(x,\delta x) \in \mathcal{M}\times T\mathcal{M}$ space obtained by adjoining a d-dimensional tangent space $\delta x \in T\mathcal{M}$ to the d-dimensional phase space $x\in\mathcal{M}\subset\mathbb{R}^d$. The matrix of variations

$$\mathbf{A}_{ij}(x) = \frac{\partial v_i(x)}{\partial x_j} \tag{4.3}$$

describes the instantaneous rate of shearing of the infinitesimal neighborhood of x by the flow. Its eigenvalues and eigendirections determine the local behavior of neighboring trajectories; nearby trajectories separate along the *unstable directions*, approach each other along the *stable directions*, and maintain their distance along the *marginal directions*. In the mathematical literature the word *neutral* is often used instead of "marginal".

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

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

one finds that the linearized neighborhood is transported by the *Jacobian* (or *fundamental*) *matrix*

$$\delta x(t) = \mathbf{J}^t(x_0)\delta x(0)\,, \qquad \mathbf{J}^t_{ij}(x_0) = \left.\frac{\partial x_i(t)}{\partial x_j}\right|_{x=x_0}\,. \tag{4.5}$$

The deformation of a neighborhood for finite time t is described by the eigenvectors and eigenvalues of the Jacobian matrix of the linearized flow. For example, consider two points along the periodic orbits separated by infinitesimal flight time δt: $\delta x(0) = f^{\delta t}(x_0) - x_0 = v(x_0)\delta t$. Time t later

$$\delta x(t) = f^{t+\delta t}(x_0) - f^t(x_0) = f^{\delta t}(x(t)) - x(t) = v(x(t))\,\delta t\,,$$

hence $\mathbf{J}^t(x_0)$ transports the velocity vector at x_0 to the velocity vector at $x(t)$ time t later:

$$v(x(t)) = \mathbf{J}^t(x_0)\, v(x_0)\,. \tag{4.6}$$

As $\mathbf{J}^t(x_0)$ eigenvalues have invariant meaning only for periodic orbits, we shall postpone discussing this to sect. 4.7.

What kinds of flows might exist? If a flow is smooth, in a sufficiently small neighborhood it is essentially linear. Hence the next section, which might seem an embarassment (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.

4.2 Linear flows

Linear fields are the simplest of vector fields. They lead to linear differential equations which can be solved explicitly, with solutions which are good for all times. The phase space for linear differential equations is $\mathcal{M} = \mathbb{R}^d$, and the differential equation (2.6) is written in terms of a vector x and a constant matrix $\mathbf{A}$ as

$$\dot{x} = v(x) = \mathbf{A}x\,. \tag{4.7}$$

Solving this equation means finding the phase 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 $x(t)'$ another solution with $x(0)' = x_0'$, then the linear combination $ax(t)+bx(t)'$ with $a, b \in \mathbb{R}$ is also a solution, but now starting at the point $ax_0 + bx_0'$. 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 (4.7)? If instead of a matrix equation we have a scalar one, $\dot{x} = ax$, with a a real number, then the solution is

$$x(t) = e^{ta}x(0)\,, \tag{4.8}$$

as you can verify by differentiation. In order to solve the matrix case, it is helpful to rederive the solution (4.8) by studying what happens for a short time step Δt. If at time 0 the position is $x(0)$, then

$$\frac{x(0+\Delta t)-x(0)}{\Delta t} = ax(0)\,, \tag{4.9}$$

which we iterate m times to obtain

$$x(t) \approx \left(1+\frac{t}{m}a\right)^m x(0)\,. \tag{4.10}$$

The term in the parenthesis 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 the parenthesis converges to e^{ta}. Consider now the matrix version of equation (4.9):

$$\frac{x(\Delta t)-x(0)}{\Delta t} = \mathbf{A}x(0)\,. \tag{4.11}$$

Representative point x is now a vector in $\mathbb{R}^d$ acted on by the matrix $\mathbf{A}$, as in (4.7). Denoting by $\mathbf{1}$ the identity matrix, and repeating the steps (4.9) and (4.10) we obtain the Euler formula for exponential of a matrix

$$x(t) = \lim_{m\to\infty}\left(\mathbf{1}+\frac{t}{m}\mathbf{A}\right)^m x(0) = e^{t\mathbf{A}}x(0)\,. \tag{4.12}$$

We will use this expression as the definition of the exponential of a matrix.

4.2.1 Operator norms

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

$$\mathbf{E}_n = \prod_{0\le m<n}\left(\mathbf{1}+\frac{t}{n}\mathbf{A}\right)$$

can be verified using the Cauchy criterion, which states that the sequence $\{\mathbf{E}_n\}$ converges if the differences $\|\mathbf{E}_k - \mathbf{E}_j\| \to 0$ 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 $\hat{n}$:

$$\|\mathbf{M}\| = \sup_{\hat{n}} \|\mathbf{M}\hat{n}\|\,, \qquad \|\hat{n}\| = 1\,. \tag{4.13}$$

We say that $\|\cdot\|$ is the operator norm induced by the vector norm $\|\cdot\|$. Constructing a norm for a finite-dimensional matrix is easy, but had $\mathbf{M}$ been an operator in an infinite-dimensional space, we would also have to specify the space $\hat{n}$ belongs to. In the finite-dimensional case, the sum of the absolute values of the components of a vector is also a norm; the induced operator norm for a matrix $\mathbf{M}$ with components M_{ij} in that case can be defined by

$$\|\mathbf{M}\| = \max_i \sum_j |M_{ij}|\,. \tag{4.14}$$

For infinite-dimensional vectors - functions $f(x), x \in \mathbb{R}^d$ - one might use instead

$$L_1 norm: \quad \int dx|f(x)|\,, \quad \text{or} l_2 norm: \quad \int dx|f(x)|^2\,,\,, etc..$$

The operator norm (4.14) and the vector norm (4.13) are only rarely distinguished by different notation, a bit of notational laziness that we shall uphold.

Now that we have learned how to make sense out of norms of operators, we can check that

$$\|e^{t\mathbf{A}}\| \leq e^{t\|\mathbf{A}\|}\,. \tag{4.15}$$

2.10 on p. 54

As $\|\mathbf{A}\|$ is a number, the norm of $e^{t\mathbf{A}}$ is finite and therefore well defined. In particular, the exponential of a matrix is well defined for all values of t, and the linear differential equation (4.7) has a solution for all times.

4.2.2 Stability eigenvalues

How do we compute the exponential (4.12)? Should we be so lucky that $\mathbf{A}$ happens to be a diagonal matrix $\mathbf{A}_D$ with eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_d)$, the exponential is simply

$$e^{t\mathbf{A}_D} = \begin{pmatrix} e^{t\lambda_1} & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & e^{t\lambda_d} \end{pmatrix}\,. \tag{4.16}$$

Usually $\mathbf{A}$ is not diagonal. In that case $\mathbf{A}$ can either be diagonalized and things are simple once again, or we have to resort to the upper triangular Jordan form.

If a matrix is a normal matrix, that is a matrix that comutes with its hermitian conjugate (the complex conjugate of its transpose), it can be diagonalized by a unitary transformation. Suppose that $\mathbf{A}$ is diagonalizable and that $\mathbf{U}$ is the matrix that brings it to its diagonal form $\mathbf{A}_D = \mathbf{U}\mathbf{A}\mathbf{U}^{-1}$. The transformation $\mathbf{U}$ is a linear coordinate transformation which rotates, skews, and possibly flips the coordinate axis of the vector space. The relation

$$e^{t\mathbf{A}} = \mathbf{U}^{-1} e^{t\mathbf{A}_D} \mathbf{U} \tag{4.17}$$

can be verified by noting that the defining product (4.10) can be rewritten as

$$\begin{aligned} e^{t\mathbf{A}} &= \left(\mathbf{U}\mathbf{U}^{-1} + \frac{t\mathbf{U}\mathbf{A}_D\mathbf{U}^{-1}}{m}\right)\left(\mathbf{U}\mathbf{U}^{-1} + \frac{t\mathbf{U}\mathbf{A}_D\mathbf{U}^{-1}}{m}\right)\cdots \\ &= \mathbf{U}\left(I + \frac{t\mathbf{A}_D}{m}\right)\mathbf{U}^{-1}\mathbf{U}\left(I + \frac{t\mathbf{A}_D}{m}\right)\mathbf{U}^{-1}\cdots = \mathbf{U}e^{t\mathbf{A}_D}\mathbf{U}^{-1}\,. \end{aligned} \tag{4.18}$$

In general, $\mathbf{A}$ will have complex eigenvalues and $\mathbf{U}$ will have complex matrix elements. The presence of complex numbers should intrigue you because in the definition of the exponential of a matrix we only used real operations. Where did the complex numbers come from?

4.2.3 Complex stability eigenvalues

As we cannot avoid complex numbers, we embrace them, and use the linearity of the vector field $\mathbf{A}x$ to extend the problem to vectors in $\mathbb{C}^d$, work there, and see the effect it has on solutions that start in $\mathbb{R}^d$. Take two vectors x and y of the phase space $\mathbb{R}^d$, combine them in a vector $w = x + iy$ in $\mathbb{C}^d$, and then extend the action of $\mathbf{A}$ to these complex vectors by $\mathbf{A}w = \mathbf{A}x + i\mathbf{A}y$. The solution $w(t)$ to the complex equation

$$\dot{w} = \mathbf{A}w \tag{4.19}$$

is the sum of the solutions $x(t) = \mathrm{Re}\,(w(t))$ and $y(t) = \mathrm{Im}\,(w(t))$ to the problem (4.7) over the reals.

To develop some intuition we work out the behavior for systems were $\mathbf{A}$ is a [2×2] matrix

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \tag{4.20}$$

The eigenvalues λ_1, λ_2 are the roots

$$\lambda_{1,2} = \frac{1}{2}\left(\operatorname{tr}\mathbf{A} \pm \sqrt{(\operatorname{tr}\mathbf{A})^2 - 4\det\mathbf{A}}\right) \tag{4.21}$$

of the characteristic equation

$$\begin{aligned} \det(\mathbf{A} - z\mathbf{1}) &= (\lambda_1 - z)(\lambda_2 - z) = 0\,, \\ \left\| \begin{matrix} A_{11} - z & A_{12} \\ A_{21} & A_{22} - z \end{matrix} \right\| &= z^2 - (A_{11} + A_{22})\,z + (A_{11}A_{22} - A_{12}A_{21}) \end{aligned} \tag{4.22}$$

The qualitative behavior of the exponential of $\mathbf{A}$ for the case that the eigenvalues λ_1 and λ_2 are both real, $\lambda_1, \lambda_2 \in \mathbb{R}$ will differ from the case that they form a complex conjugate pair, $\gamma_1, \gamma_2 \in \mathbb{C}$, $\gamma_1^* = \gamma_2$. These two possibilities are refined into sub-cases depending on the signs of the real part. The matrix might have only one linearly independent vector (an example is given sect. 5.2.1), but in general it has two linearly independent eigenvectors, which may or may not be orthogonal. Along each of these directions the motion is of the form $\exp(t\lambda_i)x_i$, $i = 1, 2$. If the eigenvalue λ_i has a positive real part, then the component x_i will grow; if the real part of λ_i is negative, it will shrink. The imaginary part of the eigenvalue leads to magnitude oscillations of x_i.

We sketch the full set of possibilities in fig. 4.1(a), and work out in detail only the case when $\mathbf{A}$ can be diagonalized in a coordinate system where the solution (4.12) to the differential equation (4.19) can be written as

$$\begin{pmatrix} w_1(t) \\ w_2(t) \end{pmatrix} = \begin{pmatrix} e^{t\lambda_1} & 0 \\ 0 & e^{t\lambda_2} \end{pmatrix} \begin{pmatrix} w_1(0) \\ w_2(0) \end{pmatrix}\,. \tag{4.23}$$

In the case $\operatorname{Re}\lambda_1 > 0$, $\operatorname{Re}\lambda_2 < 0$, w_1 grows exponentially towards infinity, and w_2 contracts towards zero. Now this growth factor is acting on the complex version of the vector, and if we want a solution to the original problem we have to concentrate on either the real or the imaginary part. The effect of the growth factor is then to make the real part of z_1 diverge to $+\infty$ if the $\operatorname{Re}(z_1) > 0$ and to $-\infty$ if the $\operatorname{Re}(z_1) < 0$. The effect on the real part of z_2 is to take it to zero. This behavior, called a *saddle*, is sketched in fig. 4.1(b), as are the remaining possibilities: in/out nodes, inward/outward spirals, and the center. saddle

Now that we have a good grip on the linear case, we are ready to return to nonlinear flows.

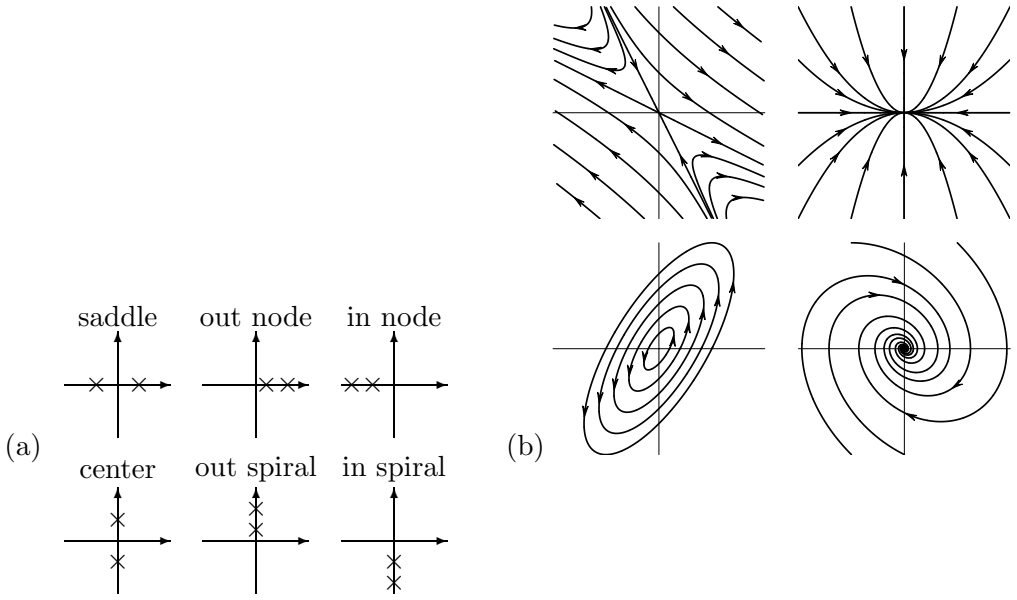


Figure 4.1: (a) Qualitatively distinct types of eigenvalues of a [2×2] stability matrix. (b) Streamlines for 2-dimensional flows.

4.3 Nonlinear flows

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

$$x(t) = f^t(x_0)\,, \quad \delta x(x_0,t) = \mathbf{J}^t(x_0)\delta x(x_0,0)\,. \tag{4.24}$$

The equations of variations are linear, so the Jacobian matrix is formally given by the integral

$$\mathbf{J}^t_{ij}(x_0) = \left[\mathbf{T}e^{\int_0^t d\tau \mathbf{A}(x(\tau))}\right]_{ij}\,. \tag{4.25}$$

appendix G.1 where $\mathbf{T}$ stands for time-ordered integration.

Let us make sense of the exponential in (4.25). For start, consider the case where x_q is an equilibrium point (2.8). Expanding around the equilibrium point x_q, using the fact that the matrix $\mathbf{A} = \mathbf{A}(x_q)$ in (4.2) is constant, and integrating,

$$f^t(x) = x_q + e^{\mathbf{A}t}(x - x_q) + \cdots\,, \tag{4.26}$$

we verify that the simple formula (4.12) applies also to the Jacobian matrix of an equilibrium point, $\mathbf{J}^t(x_q) = e^{\mathbf{A}t}$.

Next, consider the case of an arbitrary 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 (4.12):

☞ appendix J.1

$$e^{t\mathbf{A}} = \sum_{k=0}^{\infty} \frac{t^k}{k!}\mathbf{A}^k \tag{4.27}$$

$$= \lim_{m\to\infty} \left(\mathbf{1} + \frac{t}{m}\mathbf{A}\right)^m \tag{4.28}$$

Taylor expansion is fine if $\mathbf{A}$ is a constant matrix. However, only the second, tax-accountant's discrete step definition of exponential is appropriate for the task at hand, as for a dynamical system the local rate of neighborhood distortion $\mathbf{A}(x)$ depends on where we are along the trajectory. The m discrete time steps approximation to $\mathbf{J}^t$ is therefore given by generalization of the Euler product (4.12) to

$$\mathbf{J}^t = \lim_{m\to\infty} \prod_{n=m}^{1} \left(\mathbf{1} + \Delta t \mathbf{A}(x_n)\right) , \quad \Delta t = \frac{t-t_0}{m} , \quad x_n = x(t_0 + n\Delta t) , \tag{4.29}$$

with the linearized neighborhood multiplicatively deformed along the flow. To the leading order in Δt this is the same as multiplying exponentials $e^{\Delta t\, \mathbf{A}(x_n)}$, with the time ordered integral (4.25) defined as the $N \to \infty$ limit of this procedure. We note that due to the time-ordered product structure the finite time Jacobian matrices are multiplicative along the flow,

☞ appendix D

$$\mathbf{J}^{t+t'}(x_0) = \mathbf{J}^{t'}(x(t))\mathbf{J}^t(x_0) . \tag{4.30}$$

In practice, better numerical accuracy is obtained by the following observation. To linear order in Δt, $\mathbf{J}^{t+\Delta t} - \mathbf{J}^t$ equals $\Delta t\, \mathbf{A}(x(t))\mathbf{J}^t$, so the Jacobian matrix itself satisfies the linearized equation (4.1)

$$\frac{d}{dt}\mathbf{J}^t(x) = \mathbf{A}(x)\,\mathbf{J}^t(x) , \qquad \text{with the initial condition } \mathbf{J}^0(x) = \mathbf{1} . \tag{4.31}$$

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

We shall refer to the determinant $\det \mathbf{J}^t(x_0)$ as the *Jacobian* of the flow. The Jacobian is given by integral

$$\det \mathbf{J}^t(x_0) = e^{\int_0^t d\tau \mathrm{tr}\, \mathbf{A}(x(\tau))} = e^{\int_0^t d\tau \partial_i v_i(x(\tau))} . \tag{4.32}$$

This follows by computing $\det \mathbf{J}^t$ from (4.29) to the leading order in Δt. As the divergence $\partial_i v_i$ is a scalar quantity, this integral needs no time ordering. If $\partial_i v_i < 0$, the flow is *contracting*. If $\partial_i v_i = 0$, the flow preserves phase space volume and $\det \mathbf{J}^t = \mathbf{1}$. A flow with this property is called incompressible. An important class of such flows are the Hamiltonian flows to which we turn next.

in depth:
appendix J.1, p. 679

4.4 Hamiltonian flows

As the Hamiltonian flows are so important in physical applications, we digress here to illustrate the ways in which an invariance of equations of motion affects dynamics. In case at hand the *symplectic invariance* will reduce the number of independent stability exponents by factor 2 or 4.

The equations of motion for a time independent D-degrees of freedom, Hamiltonian (2.13) can be written as

$$\dot{x}_m = \omega_{mn} \frac{\partial H}{\partial x_n}, \qquad \omega = \begin{pmatrix} 0 & -\mathbf{I} \\ \mathbf{I} & 0 \end{pmatrix}, \qquad m, n = 1, 2, \ldots, 2D\,, \tag{4.33}$$

where $x = [p, q]$ is a phase space point, $\mathbf{I} = [D \times D]$ unit matrix, and ω the $[2D \times 2D]$ *symplectic* form

$$\omega_{mn} = -\omega_{nm}\,, \qquad \omega^2 = -\mathbf{1}\,. \tag{4.34}$$

The linearized motion in the vicinity $x + \delta x$ of a phase space trajectory $x(t) = (p(t), q(t))$ is described by the Jacobian matrix (4.24). The matrix of variations in (4.31) takes form

$$\mathbf{A}(x)_{mn} = \omega_{mk} H_{kn}(x)\,, \qquad \frac{d}{dt}\mathbf{J}^t(x) = \mathbf{A}(x)\mathbf{J}^t(x)\,, \tag{4.35}$$

where $H_{kn} = \partial_k \partial_n H$ is the Hessian matrix of second derivatives. From (4.35) and the symmetry of H_{kn} it follows that

$$\mathbf{A}^T \omega + \omega \mathbf{A} = 0\,. \tag{4.36}$$

This is the defining property for infinitesimal generators of *symplectic* (or canonical) transformations, transformations that leave the symplectic form ω_{mn} invariant. From this it follows that for Hamiltonian flows $\frac{d}{dt}\left(\mathbf{J}^T\omega\mathbf{J}\right) = 0$, and that $\mathbf{J}$ is a symplectic transformation (we suppress the dependence on the time and initial point, $\mathbf{J} = \mathbf{J}^t(x_0)$, $\Lambda = \Lambda(x_0, t)$, for notational brevity):

$$\mathbf{J}^T\omega\mathbf{J} = \omega\,. \tag{4.37}$$

The transpose $\mathbf{J}^T$ and the inverse $\mathbf{J}^{-1}$ are related by

$$\mathbf{J}^{-1} = -\omega\mathbf{J}^T\omega\,, \tag{4.38}$$

hence if Λ is an eigenvalue of $\mathbf{J}$, so are $1/\Lambda$, Λ^* and $1/\Lambda^*$. Real (non-marginal) eigenvalues always come paired as Λ, $1/\Lambda$. The complex eigenvalues come in pairs Λ, Λ^*, $|\Lambda| = 1$, or in loxodromic quartets Λ, $1/\Lambda$, Λ^* and $1/\Lambda^*$. Hence

4.7 on p. 96

$$\det \mathbf{J}^t(x_0) = 1 \qquad \text{for all } t \text{ and } x_0\text{'s}\,, \tag{4.39}$$

and symplectic flows preserve the *Liouville phase space volume*.

In the 2-dimensional case the eigenvalues (4.59) depend only on $\mathrm{tr}\,\mathbf{J}^t$

$$\Lambda_{1,2} = \frac{1}{2}\left(\mathrm{tr}\,\mathbf{J}^t \pm \sqrt{(\mathrm{tr}\,\mathbf{J}^t - 2)(\mathrm{tr}\,\mathbf{J}^t + 2)}\right)\,. \tag{4.40}$$

The trajectory is elliptic if the *residue* $|\mathrm{tr}\,\mathbf{J}^t| - 2 \leq 0$, with complex eigenvalues $\Lambda_1 = e^{i\theta t}$, $\Lambda_2 = \Lambda_1^* = e^{-i\theta t}$. If $|\mathrm{tr}\,\mathbf{J}^t| - 2 > 0$, the trajectory is ($\lambda$ real)

$$\text{either} \quad \textit{hyperbolic} \qquad \Lambda_1 = e^{\lambda t}\,, \quad \Lambda_2 = e^{-\lambda t}\,, \tag{4.41}$$

$$\text{or } \textit{inverse hyperbolic} \qquad \Lambda_1 = -e^{\lambda t}\,, \quad \Lambda_2 = -e^{-\lambda t}\,. \tag{4.42}$$

in depth:
appendix C.1, p. 611

4.5 Billiards

We turn next to the question of local stability of discrete time systems. Infinitesimal equations of variations (4.2) do not apply, but the multiplicative structure

Figure 4.2: Variations in the phase space coordinates of a pinball between the $(k-1)$th and the kth collision. (a) δq_k variation away from the direction of the flow. (b) δz_k angular variation tranverse to the direction of the flow. (c) $\delta q_{\|}$ variation in the direction of the flow is conserved by the flow.

(4.30) of the finite-time Jacobian matrices does. As they are more physical than most maps studied by dynamicists, let us turn to the case of billiards first.

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 the local frame of motion leads to a [2×2] Jacobian 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 kth collision of the pinball with the billiard boundary, and $t_k^{\pm} = t_k \pm \epsilon$, ϵ positive and infinitesimal. With the mass and the velocity equal to 1, the momentum direction can be specified by angle θ: $x = (q_1, q_2, \sin\theta, \cos\theta)$. Now parametrize the 2-$d$ neighborhood of a trajectory segment by $\delta x = (\delta z, \delta\theta)$, where

$$\delta z = \delta q_1 \cos\theta - \delta q_2 \sin\theta\,, \tag{4.43}$$

$\delta\theta$ is the variation in the direction of the pinball. 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 kth 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

$$\frac{d}{dt}\delta\theta = 0, \quad \frac{d}{dt}\delta z = \delta\theta\,. \tag{4.44}$$

Let $\delta\theta_k = \delta\theta(t_k^+)$ and $\delta z_k = \delta z(t_k^+)$ be the local coordinates immediately after the kth 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

$$\begin{aligned} \delta z_k^- &= \delta z_{k-1} + \tau_k \delta\theta_{k-1}\,, \qquad \tau_k = t_k - t_{k-1} \\ \delta\theta_k^- &= \delta\theta_{k-1}\,, \end{aligned} \tag{4.45}$$

and the stability matrix (4.25) for the kth free flight segment is

$$\mathbf{J}_T(x_k) = \begin{pmatrix} 1 & \tau_k \\ 0 & 1 \end{pmatrix}\,. \tag{4.46}$$

At incidence angle ϕ_k (the angle between the outgoing particle and the outgoing normal to the billiard edge), the incoming transverse variation δz_k^- projects onto an arc on the billiard boundary of length $\delta z_k^-/\cos\phi_k$. The corresponding incidence angle variation $\delta\phi_k = \delta z_k^-/\rho_k\cos\phi_k$, ρ_k = local radius of curvature, increases the angular spread to

$$\begin{aligned} \delta z_k &= -\delta z_k^- \\ \delta\theta_k &= -\delta\theta_k^- - \frac{2}{\rho_k\cos\phi_k}\delta z_k^-\,, \end{aligned} \tag{4.47}$$

so the Jacobian matrix associated with the reflection is

$$\mathbf{J}_R(x_k) = -\begin{pmatrix} 1 & 0 \\ r_k & 1 \end{pmatrix}\,, \qquad r_k = \frac{2}{\rho_k\cos\phi_k}\,. \tag{4.48}$$

The full Jacobian matrix for n_p consecutive bounces describes a beam of trajectories defocused by $\mathbf{J}_T$ along the free flight (the τ_k terms below) and defocused/refocused at reflections by $\mathbf{J}_R$ (the r_k terms below)

$$\mathbf{J}_p = (-1)^{n_p}\prod_{k=n_p}^{1}\begin{pmatrix} 1 & \tau_k \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ r_k & 1 \end{pmatrix}\,, \tag{4.49}$$

4.3 on p. 95

where τ_k is the flight time of the kth free-flight segment of the orbit, $r_k = 2/\rho_k\cos\phi_k$ is the defocusing due to the kth reflection, and ρ_k is the radius of curvature of the billiard boundary at the kth scattering point (for our 3-disk game of pinball, $\rho = 1$). As the billiard dynamics is phase-space volume preserving, $\det\mathbf{J} = 1$ and the eigenvalues are given by (4.40).

This is en example of the Jacobian matrix chain rule for discrete time systems. Stability of every flight segment or reflection taken alone is a shear with two unit

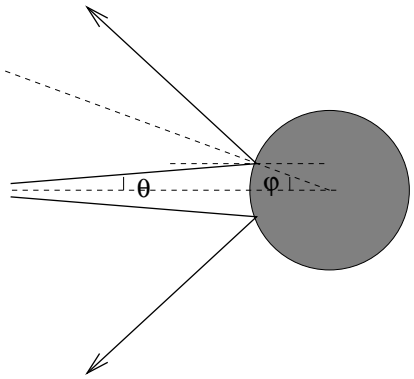


Figure 4.3: Defocusing of a beam of nearby trajectories at a billiard collision. (A. Wirzba)

eigenvalues, but acting in concert in the intervowen sequence (4.49) they can lead to a hyperbolic deformation of the infinitesimal neighborhood of a billiard trajectory.

4.4 on p. 95

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.

4.5 on p. 95
4.2 on p. 94
chapter 12

4.6 Maps

Transformation of an infinitesimal neighborhood of a trajectory under map iteration follows from Taylor expanding the iterated mapping at *discrete* time n to linear order, as in (4.4). The linearized neighborhood is transported by the Jacobian matrix

$$\mathbf{J}^n_{ij}(x_0) = \left.\frac{\partial f^n_i(x)}{\partial x_j}\right|_{x=x_0} . \tag{4.50}$$

This matrix is in the literature sometimes called the fundamental matrix. As the simplest example, a 1-dimensional map. The chain rule yields stability of the nth iterate

$$\Lambda_n = \frac{d}{dx} f^n(x) = \prod_{m=0}^{n-1} f'(x^{(m)}) , \quad x^{(m)} = f^m(x_0) . \tag{4.51}$$

The 1-step product formula for the stability of the nth iterate of a d-dimensional map

$$\mathbf{J}^n(x_0) = \prod_{m=n-1}^{0} \mathbf{J}(x^{(m)}) , \quad J_{kl}(x) = \frac{\partial}{\partial x_l} f_k(x) , \quad x^{(m)} = f^m(x_0) \tag{4.52}$$

follows from the chain rule for matrix derivatives

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

The $[d\times d]$ Jacobian matrix $\mathbf{J}^n$ for a map is evaluated along the n points $x^{(m)}$ on the trajectory of x_0, with $\mathbf{J}(x)$ the single time step Jacobian matrix. For example, for the Hénon map (3.8) the Jacobian matrix for nth iterate of the map is

4.1 on p. 94

$$\mathbf{J}^n(x_0) = \prod_{m=n}^{1} \begin{pmatrix} -2ax_m & b \\ 1 & 0 \end{pmatrix}\,, \qquad x_m = f_1^m(x_0, y_0)\,. \tag{4.53}$$

The billiard Jacobian matrix (4.49) exhibits similar multiplicative structure. The determinant of the Hénon Jacobian (4.53) is constant,

$$\det \mathbf{J} = \Lambda_1\Lambda_2 = -b \tag{4.54}$$

so in this case only one eigenvalue needs to be determined.

4.7 Cycle stabilities are metric invariants

As noted on page 35, a trajectory can be stationary, periodic or aperiodic. For chaotic systems almost all trajectories are aperiodic – nevertheless, the stationary and the 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 is a fixed point in any coordinate choice, and similarly a periodic orbit is a periodic orbit 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 mere existence of periodic orbits suffices to partially organize the spatial layout of a non–wandering set. More importantly still, as we shall now show, cycle stability eigenvalues are *metric* invariants: they determine the relative sizes of neighborhoods in a non–wandering set.

First we note that due to the multiplicative structure (4.30) of Jacobian matrices the stability of the rth repeat of a prime cycle of period T_p is

$$\mathbf{J}^{rT_p}(x_0) = \mathbf{J}^{T_p}(f^{rT_p}(x_0)) \cdots \mathbf{J}^{T_p}(f^{T_p}(x_0))\mathbf{J}^{T_p}(x_0) = \mathbf{J}_p(x_0)^r\,, \tag{4.55}$$

where $\mathbf{J}_p(x_0) = \mathbf{J}^{T_p}(x_0)$, x_0 is any point on the cycle, and $f^{rT_p}(x_0) = x_0$ by the periodicity assumption. Hence it suffices to restrict our considerations to the stability of the prime 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 (4.51) for stability of the n_pth iterate of the map

$$\Lambda_p = \frac{d}{dx} f^{n_p}(x_0) = \prod_{m=0}^{n_p-1} f'(x_m)\,, \quad x_m = f^m(x_0)\,, \tag{4.56}$$

where the initial x_0 can be any of the periodic points in the p cycle.

For future reference we note that a periodic orbit of a 1-dimensional map is *stable* if

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

and *superstable* if the orbit includes a critical point, so that the above product vanishes. A *critical point* x_c is a value of x for which the mapping $f(x)$ has vanishing derivative, $f'(x_c) = 0$. For a stable periodic orbit of period n the slope of the nth iterate $f^n(x)$ evaluated on a periodic point x (fixed point of the nth iterate) lies between -1 and 1.

The 1-dimensional map (4.51) cycle stability Λ_p is a product of derivatives over all cycle points around the cycle, and is therefore independent of which periodic point is chosen as the initial one. In higher dimensions the Jacobian matrix $\mathbf{J}_p(x_0)$ in (4.55) does depend on the initial point $x_0 \in p$. However, as we shall now show, the cycle stability eigenvalues are intrinsic property of a cycle in any dimension. Consider the ith eigenvalue, eigenvector evaluated at a cycle point x,

$$\mathbf{J}_p(x)\mathbf{e}^{(i)}(x) = \Lambda_{p,i}\mathbf{e}^{(i)}(x)\,, \quad x \in p\,,$$

and at another point on the cycle $x' = f^t(x)$. By the chain rule (4.30) the Jacobian matrix at x' can be written as

$$\mathbf{J}^{T_p+t}(x) = \mathbf{J}^{T_p}(x')\mathbf{J}^t(x) = \mathbf{J}_p(x')\mathbf{J}^t(x).$$

Defining the eigenvactor transported along the flow $x \to x'$ by $\mathbf{e}^{(i)}(x') = \mathbf{J}^t(x)\mathbf{e}^{(i)}(x)$, we see that $\mathbf{J}_p$ evaluated anywhere along the cycle has the same set of stability eigenvalues $\{\Lambda_{p,1}, \Lambda_{p,2}, \cdots \Lambda_{p,d}\}$

$$\left(\mathbf{J}_p(x') - \Lambda_{p,i}\mathbf{1}\right)\mathbf{e}^{(i)}(x') = 0\,, \quad x' \in p\,. \tag{4.57}$$

Quantities such as $\mathrm{tr}\,\mathbf{J}_p(x)$, $\det\mathbf{J}_p(x)$ depend only on the eigenvalues of $\mathbf{J}_p(x)$ and not on x, hence in expressions such as

$$\det\left(\mathbf{1}-\mathbf{J}_p^r\right)=\det\left(\mathbf{1}-\mathbf{J}_p^r(x)\right) \tag{4.58}$$

we will omit reference to any particular cycle point x.

We sort the *stability eigenvalues* $\Lambda_{p,1}$, $\Lambda_{p,2}$, ..., $\Lambda_{p,d}$ of the $[d\times d]$ Jacobian matrix $\mathbf{J}_p$ evaluated on the p cycle into sets $\{e,m,c\}$

$$\begin{aligned}
\text{expanding:}\quad & \{\Lambda_{p,e}\} &&= \{\Lambda_{p,i}:|\Lambda_{p,i}|>1\}\\
\text{marginal:}\quad & \{\Lambda_{p,m}\} &&= \{\Lambda_{p,i}:|\Lambda_{p,i}|=1\}\\
\text{contracting:}\quad & \{\Lambda_{p,c}\} &&= \{\Lambda_{p,i}:|\Lambda_{p,i}|<1\}\,.
\end{aligned} \tag{4.59}$$

and denote by Λ_p (no spatial index) the product of expanding eigenvalues

$$\Lambda_p=\prod_e \Lambda_{p,e}\,. \tag{4.60}$$

Cycle *stability exponents* are defined as (see (4.16) (4.41) and (4.42) for examples) as stretching/contraction rates per unit time

$$\lambda_{p,i}=\ln|\Lambda_{p,i}|\,T_p \tag{4.61}$$

We distinguish three cases

$$\begin{aligned}
\text{expanding:}\quad & \{\lambda_{p,e}\} &&= \{\lambda_{p,e}:\ \lambda_{p,e}>0\}\\
\text{elliptic:}\quad & \{\lambda_{p,m}\} &&= \{\lambda_{p,m}:\ \lambda_{p,m}=0\}\\
\text{contracting:}\quad & \{\lambda_{p,c}\} &&= \{\lambda_{p,c}:\lambda_{p,c}<0\}\,.
\end{aligned} \tag{4.62}$$

Cycle stability exponents are of interest because they are a local version of the Lyapunov exponents (to be discussed below in sect. 6.3). However, we do care about the sign of $\Lambda_{p,i}$ and its phase (if $\Lambda_{p,i}$ is complex), and keeping track of those by case-by-case enumeration, as in (4.41) - (4.42), is a nuisance, so almost all of our formulas will be stated in terms of stability eigenvalues $\Lambda_{p,i}$ rather than in terms of stability exponents $\lambda_{p,i}$.

Our task will be to determine the size of a neighborhood, and that is why we care about the stability eigenvalues, 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 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 Λ_p is the product of expanding eigenvalues (4.60) only; contracting ones play a secondary role.

Presence of marginal eigenvalues signals either an invariance of the flow (which you should immediately exploit to simplify the problem), or a non-hyperbolicity of a flow (source of much pain, hard to avoid).

☞ chapter 16

A periodic orbit always has at least one marginal eigenvalue. As $\mathbf{J}^t(x)$ transports the velocity field $v(x)$ by (4.6), after a complete period

$$\mathbf{J}_p(x)v(x) = v(x)\,, \tag{4.63}$$

and a periodic orbit always has an eigenvector $\mathbf{e}^{(\|)}$ parallel to the local velocity field with eigenvalue

$$\Lambda_{p,\|} = 1\,. \tag{4.64}$$

A periodic orbit p of a d-dimensional map is *stable* if the magnitude of every one of its stability eigenvalues is less than one, $|\Lambda_{p,i}| < 1$ for $i = 1, 2, \ldots, d$. The region of parameter values for which a periodic orbit p is stable is called the *stability window* of p.

4.7.1 Smooth conjugacies

So far we have established that for a given flow the cycle stability eigenvalues are intrinsic to a given cycle. As we shall now see, they are intrinsic to the cycle in any representation of the dynamical system.

That the cycle stability eigenvalues are an invariant property of the given dynamical system follows from elementary considerations of sect. 2.3: 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 (2.17), a reparametrization 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,

$$\begin{aligned} g'(y) &= h'(f \circ h^{-1}(y)) f'(h^{-1}(y)) \frac{1}{h'(x)} \\ &= h'(x) f'(x) \frac{1}{h'(x)} = f'(x)\,, \end{aligned} \tag{4.65}$$

and the generalization to the stability eigenvalues of periodic orbits of d-dimensional flows is immediate.

As stability of a flow can always be rewritten as stability of a Poincaré section return map, we find that the stability eigenvalues of any cycle, for a flow or a map in arbitrary dimension, is a metric invariant of the dynamical system.

2.7 on p. 53

4.8 Going global: Stable/unstable manifolds

The invariance of stabilities of a periodic orbit is a local property of the flow. Now we show that every periodic orbit carries with it stable and unstable manifolds which provide a *global* topologically invariant foliation of the phase space.

The fixed or periodic point x^* stability matrix $\mathbf{J}_p(x^*)$ eigenvectors describe the flow into or out of the fixed point only infinitesimally close to the fixed point. 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

$$\begin{aligned} W^s &= \left\{x \in \mathcal{M} : f^t(x) - x^* \to 0 \text{ as } t \to \infty\right\} \\ W^u &= \left\{x \in \mathcal{M} : f^{-t}(x) - x^* \to 0 \text{ as } t \to \infty\right\}\,. \end{aligned} \tag{4.66}$$

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 parcourse, we might just as well look at their Poincaré section return maps. Stable, unstable manifolds for maps are defined by

$$\begin{aligned} W^s &= \{x \in \mathcal{P} : f^n(x) - x^* \to 0 \text{ as } n \to \infty\} \\ W^u &= \left\{x \in \mathcal{P} : f^{-n}(x) - x^* \to 0 \text{ as } n \to \infty\right\}\,. \end{aligned} \tag{4.67}$$

For $n \to \infty$ any finite segment of W^s, respectively W^u converges to the linearized map eigenvector ϵ^s, respectively ϵ^u. 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: The stable (unstable) manifold of the central hyperbolic fixed point (x_1, x_1) can be constructed numerically by starting with a small interval along the local stable (unstable) eigendirection, and iterating the interval n steps backwards (forwards).

Both in the example of the Rössler flow and of the Kuramoto-Sivashinsky system we have learned that the attractor is very thin, but otherwise the return maps that we found were disquieting – neither fig. 3.2 nor fig. 3.3 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 was entirely 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? T As we shall see in sect. **??**, 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 unstable manifold.

in depth:
appendix C.1, p. 611

Commentary

Remark **4.1** Further reading. The chapter 1 of Gaspard's monograph [4] is recommended reading if you are interested in Hamiltonian flows, and billiards in particular. A. Wirzba has generalized the stability analysis of sect. 4.5 to scattering off 3-dimensional spheres (follow the links in `www.nbi.dk/-ChaosBook/extras/`). A clear discussion of linear stability for the general d-dimensional case is given in Gaspard [4], sect. 1.4.

Résumé

A neighborhood of a trajectory deforms as it is transported by the flow. In the linear approximation, the matrix of variations $\mathbf{A}$ describes this shearing of an infinitesimal neighborhood in an infinitesimal time step. The shearing after finite time is described by the Jacobian matrix$\mathbf{J}^t$. 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.

Periodic orbits play a central role in any invariant characterization of the dynamics, as their existence and inter-relations are *topological*, coordinate choice independent property of the dynamics. Furthermore, they form an infinite set of *metric invariants*: The stability eigenvalues of a periodic orbit remain invariant under any smooth nonlinear change of coordinates $f \to h \circ f \circ h^{-1}$.

Exercises

4.1 How unstable is the Hénon attractor?

(a) Evaluate numerically the Lyapunov exponent by iterating the Hénon map

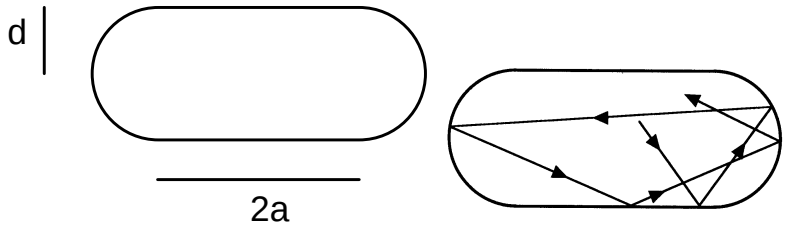
$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 - ax^2 + y \\ bx \end{bmatrix}$$

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?

4.2 A pinball simulator. Add to your exercise 3.7 pinball simulator a routine that computes the the [2×x2] Jacobian matrix. To be able to compare with the numerical results of coming chapters, work with R:$a = 6$ and/or 2.5 values.

4.3 Stadium billiard. The *Bunimovich stadium* [**?**, **?**] is a billiard with a point particle moving freely within a two dimensional domain, reflected elastically at the border which consists of 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 on the border. The length a is the only parameter.

The Jacobian matrix associated with the reflection is given by (4.48). Here we take $\rho_k = -1$ for the semicircle sections of the boundary, and $\cos\phi_k$ remains constant for all

bounces in a rotation sequence. The time of flight between two semicircle bounces is $\tau_k = 2\cos\phi_k$. The Jacobian matrix of one semicircle reflection folowed by the flight to the next bounce is

$$\mathbf{J} = (-1)\begin{pmatrix} 1 & 2\cos\phi_k \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -2/\cos\phi_k & 1 \end{pmatrix} = (-1)\begin{pmatrix} -3 & 2\cos\phi_k \\ 2/\cos\phi_k & 1 \end{pmatrix}.$$

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 Jacobian matrix given by shear (*ie.* the eigenvalues remain equal to 1 throughout the whole rotation), and k bounces inside a circle lead to

$$\mathbf{J}^k = (-1)^k\begin{pmatrix} -2k-1 & 2k\cos\phi \\ 2k/\cos\phi & 2k-1 \end{pmatrix}. \tag{4.68}$$

The Jacobian matrix of a cycle p of length n_p is given by

$$\mathbf{J}_p = (-1)^{\sum n_k}\prod_{k=1}^{n_p}\begin{pmatrix} 1 & \tau_k \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ n_k r_k & 1 \end{pmatrix}. \tag{4.69}$$

Adopt your pinball simulator to the Bunimovich stadium.

4.4 Fundamental domain fixed points. Use the formula (4.49) for billiard Jacobian matrix to compute the periods T_p and the expanding eigenvalues Λ_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	Λ_p
$\overline{0}$:	$R-2$	$R-1+R\sqrt{1-2/R}$
$\overline{1}$:	$R-\sqrt{3}$	$-\frac{2R}{\sqrt{3}}+1-\frac{2R}{\sqrt{3}}\sqrt{1-\sqrt{3}/R}$

(4.70)

We have set the disk radius to $a = 1$.

4.5 Fundamental domain 2-cycle. Verify that for the $\overline{10}$-cycle the cycle length and the trace of the Jacobian matrix are given by

$$\begin{aligned} L_{10} &= 2\sqrt{R^2-\sqrt{3}R+1}-2, \\ \operatorname{tr}\mathbf{J}_{10} &= 2L_{10}+2+\frac{1}{2}\,\frac{L_{10}(L_{10}+2)^2}{\sqrt{3}R/2-1}. \end{aligned} \tag{4.71}$$

The $\overline{10}$-cycle is drawn in fig. 10.4. The unstable eigenvalue Λ_{10} follows from (4.21).

4.6 A test of your pinball simulator. Test your exercise 4.2 pinball simulator by comparing what it yields with the analytic formulas of exercise 4.4 and 4.5.

4.7 Birkhoff coordinates. Prove that the Birkhoff coordinates are phase-space volume preserving. Hint: compute the determinant of (4.49).

Chapter 5

Transporting densities

> O what is my destination? (I fear it is henceforth chaos;)
> Walt Whitman,
> *Leaves of Grass: Out of the Cradle Endlessly Rocking*

(P. Cvitanović, R. Artuso, L. Rondoni, and E.A. Spiegel)

In chapters 2 and 3 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, 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 evolution operators.

5.1 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

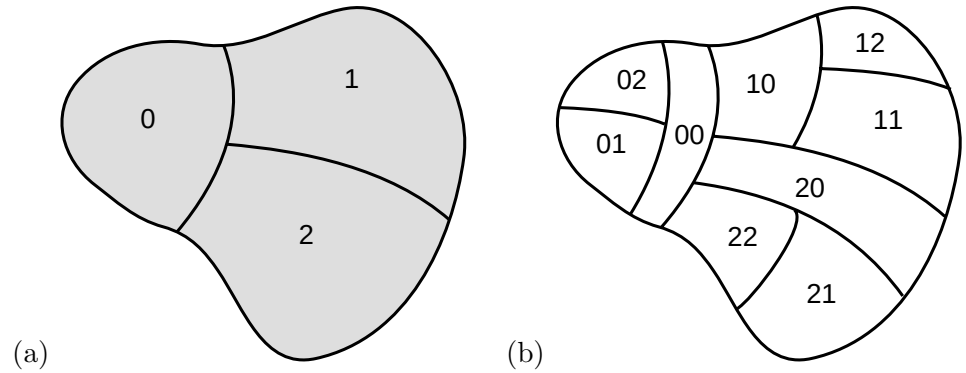


Figure 5.1: (a) First level of partitioning: A coarse partition of $\mathcal{M}$ into regions $\mathcal{M}_0$, $\mathcal{M}_1$, and $\mathcal{M}_2$. (b) $n = 2$ level of partitioning: A refinement of the above partition, with each region $\mathcal{M}_i$ subdivided into $\mathcal{M}_{i0}$, $\mathcal{M}_{i1}$, and $\mathcal{M}_{i2}$.

and construct a physically meaningful measure is by a process of coarse-graining. Consider a sequence 1, 2, ..., n, ... of more and more refined partitions of the phase space, fig. 5.1, into regions $\mathcal{M}_i$ defined by the characteristic function

$$\chi_i(x) = \begin{cases} 1 & \text{if } x \in \text{ region } \mathcal{M}_i \\ 0 & \text{otherwise} \end{cases} . \tag{5.1}$$

A coarse-grained measure is obtained by assigning the "mass", or the fraction of trajectories contained in the ith region $\mathcal{M}_i \subset \mathcal{M}$ at the nth level of partitioning of the phase space:

$$\Delta\mu_i = \int_{\mathcal{M}} d\mu(x)\chi_i(x) = \int_{\mathcal{M}_i} d\mu(x) = \int_{\mathcal{M}_i} dx\, \rho(x)\,. \tag{5.2}$$

$\rho(x) = \rho(x,t)$ is the *density* of representative points in the phase 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. As our intent is to sprinkle the phase space with a finite number of initial points, we shall assume that the measure can be normalized

$$\sum_i^{(n)} \Delta\mu_i = 1\,, \tag{5.3}$$

where the sum is over subregions i at the nth level of partitioning. The infinitesimal measure $dx\rho(x)$ can be thought of as a continuum limit of $\Delta\mu_i = |\mathcal{M}_i|\rho(x_i)\,, \quad x_i \in \mathcal{M}_i$, with normalization

$$\int_{\mathcal{M}} dx\, \rho(x) = 1\,. \tag{5.4}$$

While dynamics can lead to very singular ρ's, in practice we cannot do better than to measure it averaged over some region $\mathcal{M}_i$, and that is why we insist on "coarse-graining" here. 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 there are no probabilistic evolution kernels, the density of trajectories is transported *deterministically*. What this distinction means will became apparent later on: for deterministic flows our trace and determinant formulas will be *exact*, while for quantum and stochastic flows they will only be the leading saddlepoint approximations.

☞ chapter 8

☞ chapter ??

So far, any arbitrary sequence of partitions will do. What are intelligent ways of partitioning the phase space? We postpone the answer to chapter **??**, after we have developed some intuition about how the dynamics transports densities.

☞ chapter ??

5.2 Density evolution

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 $t = 0$. As the flow evolves, this region is carried into $f^t(\mathcal{M}_i)$, as in fig. 2.1(b). No trajectory is created or destroyed, so the conservation of representative points requires that

$$\int_{f^t(\mathcal{M}_i)} dx\, \rho(x,t) = \int_{\mathcal{M}_i} dx_0\, \rho(x_0,0)\,.$$

If the flow is invertible and the transformation $x_0 = f^{-t}(x)$ is single valued, we can transform the integration variable in the expression on the left to

$$\int_{\mathcal{M}_i} dx_0\, \rho(f^t(x_0),t)\left|\det \mathbf{J}^t(x_0)\right|\,.$$

We conclude that the density changes with time as the inverse of the Jacobian (4.32)

$$\rho(x,t) = \frac{\rho(x_0,0)}{|\det \mathbf{J}^t(x_0)|}\,, \qquad x = f^t(x_0)\,, \tag{5.5}$$

which makes sense: the density varies inversely to the infinitesimal volume occupied by the trajectories of the flow. The manner in which a flow transports densities may be recast into language of operators, by writing

$$\rho(x,t) = \mathcal{L}^t\rho(x) = \int_{\mathcal{M}} dx_0\, \delta\big(x - f^t(x_0)\big)\, \rho(x_0,0)\,. \tag{5.6}$$

Let us check this formula. Integrating Dirac delta functions is easy: $\int_{\mathcal{M}} dx\, \delta(x) = 1$ if $0 \in \mathcal{M}$, zero otherwise. Integral over a one-dimensional Dirac delta function picks up the Jacobian of its argument evaluated at all of its zeros:

$$\int dx\, \delta(h(x)) = \sum_{x \in \text{Zero}\,[h]} \frac{1}{|h(x)'|}\,, \tag{5.7}$$

5.1 on p. 112 and in d dimensions the denominator is replaced by

$$\int dx\, \delta(h(x)) = \sum_{x \in \text{Zero}\,[h]} \frac{1}{\left|\det \frac{\partial h(x)}{\partial x}\right|}\,. \tag{5.8}$$

5.2 on p. 112 Now you can check that (5.6) is just a rewrite of (5.5):

$$\begin{aligned} \mathcal{L}^t \rho(x) &= \sum_{x_0 = f^{-t}(x)} \frac{\rho(x_0)}{|f^t(x_0)'|} && \text{(1-dimensional)} \\ &= \sum_{x_0 = f^{-t}(x)} \frac{\rho(x_0)}{|\det \mathbf{J}^t(x_0)|} && (d\text{-dimensional})\,. \end{aligned} \tag{5.9}$$

For a deterministic, invertible flow there is only one x_0 preimage of x; allowing for multiple preimages also takes account of noninvertible mappings such as the "stretch&fold" maps of the interval, to be discussed in the next example, or more generally in sect. 10.5.

5.3 on p. 113 We shall refer to the kernel of (5.6) as the *Perron-Frobenius operator*:

$$\mathcal{L}^t(x,y) = \delta\big(x - f^t(y)\big)\ . \tag{5.10}$$

sect. 9.3.1

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. 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 $t=0$.

in depth:
appendix D, p. 617

5.2.1 A piecewise-linear example

What is gained by reformulation of dynamics in terms of "operators"? We start by considering a simple example where the operator is a $[2 \times 2]$ matrix. Assume

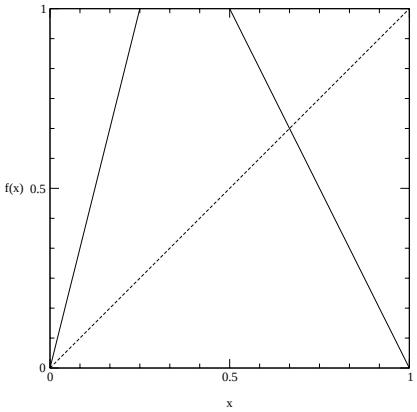


Figure 5.2: A piecewise-linear repeller: All trajectories that land in the gap between the f_0 and f_1 branches escape.

the expanding 1-d map $f(x)$ of fig. 5.2, a piecewise-linear 2–branch repeller with slopes $\Lambda_0 > 1$ and $\Lambda_1 < -1$:

$$f(x) = \begin{cases} f_0 = \Lambda_0 x & \text{if} \quad x \in \mathcal{M}_0 = [0, 1/\Lambda_0] \\ f_1 = \Lambda_1 (x-1) & \text{if} \quad x \in \mathcal{M}_1 = [1 + 1/\Lambda_1, 1] \end{cases} . \tag{5.11}$$

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

$$\rho(x) = \begin{cases} \rho_0 & \text{if } x \in \mathcal{M}_0 \\ \rho_1 & \text{if } x \in \mathcal{M}_1 \end{cases} . \tag{5.12}$$

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, fig. 1.6, with f_0 and f_1 modelling its two strips of survivors.

As can be easily checked by using (5.9), the Perron-Frobenius operator acts on this piecewise constant function as a [2×2] "transfer" matrix with matrix elements

5.5 on p. 114

$$\begin{pmatrix} \rho_0 \\ \rho_1 \end{pmatrix} \to \mathcal{L}\rho = \begin{pmatrix} \frac{1}{|\Lambda_0|} & \frac{1}{|\Lambda_1|} \\ \frac{1}{|\Lambda_0|} & \frac{1}{|\Lambda_1|} \end{pmatrix} \begin{pmatrix} \rho_0 \\ \rho_1 \end{pmatrix} , \tag{5.13}$$

stretching both ρ_0 and ρ_1 over the whole unit interval Λ, and decreasing the density at every iteration. As in this example the density is constant after one iteration, $\mathcal{L}$ has only one eigenvalue $e^{s_0} = 1/|\Lambda_0| + 1/|\Lambda_1|$, with the constant density eigenvector $\rho_0 = \rho_1$. $1/|\Lambda_0|$, $1/|\Lambda_1|$ are respectively the sizes of $|\mathcal{M}_0|$, $|\mathcal{M}_1|$ intervals, so the *exact* escape rate (1.3) – the log of the fraction of survivors at each iteration for this linear repeller – is given by the sole eigenvalue of $\mathcal{L}$:

$$\gamma = -s_0 = -\ln(1/|\Lambda_0| + 1/|\Lambda_1|) . \tag{5.14}$$

Voila! Here is the rationale for introducing operators – in one time step we have solved the problem of evaluating escape rate at infinite time. Such simple explicit matrix representation of the Perron-Frobenius operator is a consequence of piecewise linearity of f, and the restriction of the densities ρ to the space of piecewise constant functions. In general case there will exist no such finite-dimensional representation for the Perron-Frobenius operator. To a student with practical bend the example does suggest a strategy for constructing evolution operators for smooth maps, as limits of partitions of phase space into regions $\mathcal{M}_i$, with a piecewise-linear approximation f_i to dynamics in each region, but that would be too naive; much of the physically interesting spectrum would be missed. As we shall see, the choice of function space for ρ is crucial, and the physically motivated choice is a space of smooth functions, rather than the space of piecewise constant functions.

☞ chapter 9

5.3 Invariant measures

A *stationary* or *invariant density* is a density left unchanged by the flow

$$\rho(f^t(x)) = \rho(x) = \rho(f^{-t}(x))\,. \tag{5.15}$$

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 computation of asymptotic averages of observables, our task is to identify interesting invariant measures for a given $f^t(x)$. Invariant measures remain unaffected by dynamics, so they are fixed points (in the infinite-dimensional function space of ρ densities) of the Perron-Frobenius operator (5.10), with the unit eigenvalue:

5.3 on p. 113

$$\mathcal{L}^t\rho(x) = \int_{\mathcal{M}} dy\, \delta(x - f^t(y))\rho(y) = \rho(x). \tag{5.16}$$

Depending on the choice of $f^t(x)$, there may be no, one, or many solutions of the eigenfunction condition (5.16). For instance, a singular measure $d\mu(x) = \delta(x - x^*)dx$ concentrated on an equilibrium point $x^* = f^t(x^*)$, or any linear combination of such measures, each concentrated on a different equilibrium point, is stationary. So there are infinitely many stationary measures you can construct, almost all of them unnatural in the sense that a slightest perturbation will destroy them. Intutitively, the "natural" measure should be the one least sensitive to inevitable facts of life, such as noise, not matter how weak.

5.3.1 Natural measure

The *natural* or *equilibrium measure* can be defined as the limit

$$\overline{\rho}_{x_0}(y) = \lim_{t\to\infty} \frac{1}{t} \int_0^t d\tau\, \delta(y - f^\tau(x_0)) \ , \tag{5.17}$$

5.8 on p. 115
5.9 on p. 115

where x_0 is a generic inital point. Staring at an average over ∞ many Dirac deltas is not a prospect we cherish. From a physical point of view, it is more sensible to think of the natural measure as a limit of the transformations which an initial smooth distribution experiences under the action of f, rather than as a limit computed from a single trajectory. Generated by the action of f, the natural measure satisfies the stationarity condition (5.16) and is invariant by construction. From the computational point of view, the natural measure is the visitation frequency defined by coarse-graining, integrating (5.17) over the $\mathcal{M}_i$ region

$$\Delta\overline{\mu}_i = \lim_{t\to\infty} \frac{t_i}{t} \,, \tag{5.18}$$

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*, a function belonging to some function space, for instance the space of integrable functions L^1, that associates to each point in phase space a number or a set of numbers. The observable reports on some property of the dynamical system (several examples will be given in sect. 6.1). The *space average* of the observable a with respect to measure ρ is given by the d-dimensional integral over the phase space $\mathcal{M}$:

$$\langle a \rangle = \frac{1}{|\rho_{\mathcal{M}}|} \int_{\mathcal{M}} dx\, \rho(x) a(x) \,, \quad |\rho_{\mathcal{M}}| = \int_{\mathcal{M}} dx\, \rho(x) = \text{mass in } \mathcal{M} \,. \tag{5.19}$$

For the time being we assume that the phase space $\mathcal{M}$ has a finite dimension and a finite volume. By its definition $\langle a \rangle$ is a function(al) of ρ , $\langle a \rangle = \langle a \rangle_\rho$.

Inserting the right hand side of (5.17) into (5.19) we see that the natural measure corresponds to *time average* of the observable a along a trajectory of the initial point x_0,

$$\overline{a(x_0)} = \lim_{t\to\infty} \frac{1}{t} \int_0^t d\tau\, a(f^\tau(x_0)) \,. \tag{5.20}$$

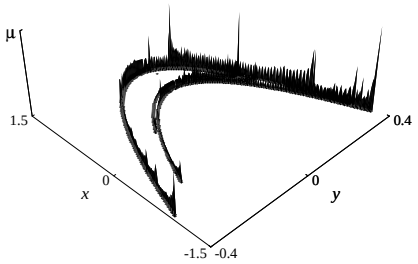


Figure 5.3: Natural measure (5.18) for the Hénon map (3.8) strange attractor at parameter values $(a, b) = (1.4, 0.3)$. See fig. 3.4 for a sketch of the attractor without the natural measure binning. (Courtesy of J.-P. Eckmann)

Analysis of the above asyptotic time limit is the central problem of ergodic theory. More precisely, the Birkhoff ergodic theorem asserts that if a natural measure ρ exists, the limit $\overline{a(x_0)}$ for the time average (5.20) 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 over almost any trajectory tends to the space average

☞ appendix B

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t d\tau\, a(f^\tau(x_0)) = \langle a \rangle \tag{5.21}$$

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 ρ-measure zero. For future reference, we note a further property, stronger than ergodicity: if you can establish the space average of a product of any two variables decorrelates with time,

$$\lim_{t\to\infty} \langle a(0)b(t) \rangle = \langle a \rangle \langle b \rangle \ , \tag{5.22}$$

☞ sect. 14.3

the dynamical system is said to be *mixing*.

An example of a numerical calculation of the natural measure (5.18) for the Hénon attractor (3.8) is given in fig. 5.3. The phase space is partitioned into many equal size areas ρ_i, and the coarse grained measure (5.18) computed by a long time iteration of the Hénon map, and represented by the height of the pin over area $\mathcal{M}_i$. What you see is a typical invariant measure complicated, singular function concentrated on a fractal set. If an invariant measure is quite singular (for instance a Dirac δ concentrated on a fixed point or a cycle), its existence is most likely of limited physical import. No smooth inital density will converge to this measure if the dynamics is unstable. In practice the average (5.17) 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, is a strange attractor or merely a long stable cycle.

4.1 on p. 94

Clearly, while deceptively easy to define, measures spell trouble. The good news is that if you hang on, you will never ever need to compute them. How

so? The evolution operators that we turn to next, and the trace and determinant formulas that they will lead us to will assign the correct natural measure weights to desired averages without recourse to any explicit computation of the coarse-grained measure $\Delta\mathcal{M}_i$.

5.4 Koopman, Perron-Frobenius operators

> 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*

The way in which time evolution acts on densities may be rephrased in the language of functional analysis, by introducing the *Koopman operator*, whose action on a phase 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 $x(t)$:

$$\mathcal{K}^t a(x) \;=\; a(f^t(x))\,. \tag{5.23}$$

Observable $a(x)$ has no explicit time dependence; all time dependence is carried in 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 (5.10) depends on the trajectory point time t in the past, so the Perron-Frobenius operator is the adjoint of the Koopman operator

5.10 on p. 115

$$\int_{\mathcal{M}} dx\;\left[\mathcal{K}^t a(x)\right]\,\rho(x) = \int_{\mathcal{M}} dx\; a(x)\;\left[\mathcal{L}^t\rho(x)\right]\,. \tag{5.24}$$

Checking this is an easy change of variables exercise. For finite dimensional deterministic invertible flows the Koopman operator (5.23) is simply the inverse of the Perron-Frobenius operator (5.6), so in what follows we shall not distinguish

sect. 2.5.3

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 parametrized by time

(a) $\mathcal{K}^0 = I$

(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

$$\mathcal{A} \;=\; \lim_{t\to 0^+} \frac{1}{t}\left(\mathcal{K}^t - I\right) .$$

(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 (2.4):

$$\mathcal{A}a(x) \;=\; \lim_{t\to 0^+} \frac{1}{t}\left(a(f^t(x)) - a(x)\right) \;=\; v_i(x)\partial_i a(x)\,. \tag{5.25}$$

Of course, that is nothing but the definition of the time derivative, so the equation of motion for $a(x)$ is

$$\left(\frac{d}{dt} - \mathcal{A}\right) a(x) \;=\; 0\,. \tag{5.26}$$

The finite time Koopman operator (5.23) can be formally expressed by exponentiating the time evolution generator $\mathcal{A}$ as

$$\mathcal{K}^t = e^{t\mathcal{A}}\,. \tag{5.27}$$

5.11 on p. 115

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:

5.12 on p. 115

$$e^{tv\frac{\partial}{\partial x}} a(x) = a(x + tv)\,. \tag{5.28}$$

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 D.2. Here we limit ourselves to a brief remark about the notion of "spectrum" of a linear operator.

The Koopman operator $\mathcal{K}$ acts multiplicatively in time, so it is reasonable to suppose that there exist constants $M > 0$, $\beta \geq 0$ such that $||\mathcal{K}^t|| \leq Me^{t\beta}$ for all $t \geq 0$. What does that mean? The operator norm is defined in the same spirit in which we defined the matrix norms in sect. 4.2.1: 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 (5.14). 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 $\mathcal{A} - \beta I$. Given this bound, it follows by the Laplace transform

$$\int_0^\infty dt\, e^{-st}\mathcal{K}^t = \frac{1}{s-\mathcal{A}}\,, \qquad \mathrm{Re}\, s > \beta\,, \tag{5.29}$$

that the *resolvent* operator $(s-\mathcal{A})^{-1}$ is bounded

$$\left|\left|\frac{1}{s-\mathcal{A}}\right|\right| \leq \int_0^\infty dt\, e^{-st}Me^{t\beta} = \frac{M}{s-\beta}\,.$$

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 (5.27) into the resolvent form (5.29) and enables us to study its spectrum.

in depth:
appendix D.2, p. 618

5.4.1 Liouville operator

A case of special interest is the Hamiltonian or symplectic flow defined by the time-independent Hamiltonian equations of motion (2.13). 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, (5.26) looks rather much 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.

For separable Hamiltonians of form $H = p^2/2m + V(q)$, the equations of motion are

$$\dot{q}_i = \frac{p_i}{m}\,, \qquad \dot{p}_i = -\frac{\partial V(q)}{\partial q_i}\,. \tag{5.30}$$

The evolution equations for any p, q dependent quantity $Q = Q(p,q)$ are given by

$$\frac{dQ}{dt} = \frac{\partial Q}{\partial q_i}\frac{dq_i}{dt} + \frac{\partial Q}{\partial p_i}\frac{dp_i}{dt} = \frac{\partial H}{\partial p_i}\frac{\partial Q}{\partial q_i} - \frac{\partial Q}{\partial p_i}\frac{\partial H}{\partial q_i}\,. \tag{5.31}$$

As equations with this structure arise frequently for symplectic flows, it is convenient to introduce a notation for them, the *Poisson bracket*

$$[A,B] = \frac{\partial A}{\partial p_i}\frac{\partial B}{\partial q_i} - \frac{\partial A}{\partial q_i}\frac{\partial B}{\partial p_i}\,. \tag{5.32}$$

In terms of Poisson brackets the time evolution equation (5.31) takes the compact form

$$\frac{dQ}{dt} = [H,Q]\,. \tag{5.33}$$

The phase space flow velocity is $\mathbf{v} = (\dot{q},\dot{p})$, where the dot signifies time derivative for fixed initial point. Hamilton's equations (2.13) imply that the flow is incompressible, $\partial_i v_i = 0$, so for Hamiltonian flows the equation for ρ reduces to the *continuity equation* for the density:

☞ appendix D

$$\partial_t \rho + \partial_i(\rho v_i) = 0\,. \tag{5.34}$$

Consider evolution of the phase space density ρ of an ensemble of noninteracting particles subject to the potential $V(q)$; 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 (2.13) we obtain the *Liouville equation*, a special case of (5.26):

$$\frac{\partial}{\partial t}\rho(q,p,t) = -\mathcal{A}\rho(q,p,t) = [H,\rho(q,p,t)]\,, \tag{5.35}$$

where [,] is the Poisson bracket (5.32). The generator of the flow (5.25) is now the generator of infinitesimal symplectic transformations,

$$\mathcal{A} = \dot{q}_i \frac{\partial}{\partial q_i} + \dot{p}_i \frac{\partial}{\partial p_i} = \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q_i} - \frac{\partial H}{partial q_i}\frac{\partial}{\partial p_i} \,. \tag{5.36}$$

or, by the Hamilton's equations for separable Hamiltonians

$$\mathcal{A} = -\frac{p_i}{m}\frac{\partial}{\partial q_i} + \partial_i V(q)\frac{\partial}{\partial p_i} \,. \tag{5.37}$$

5.13 on p. 116

This special case of the time evolution generator (5.25) for the case of symplectic flows is called the *Liouville operator*. You might have encountered it in statistical mechanics, in rather formal settings, while discussing what ergodicity means for 10^{23} hard balls, or on the road from Liouville to Boltzmann. Here its action will be very tangible; we shall apply the evolution operator to systems as small as 1 or 2 hard balls and to our suprise learn that suffices to get a grip on some of the fundations of the classical nonequilibrium statistical mechanics.

in depth: sect. D.2, p. 618

Commentary

Remark **5.1** Ergodic theory. An overview of ergodic theory is outside the scope of this book: the interested reader may find it useful to consult [1]. The existence of time average (5.20) is the basic result of ergodic theory, known as the Birkhoff theorem, see for example refs. [1, 2], or the statement of the theorem 7.3.1 in ref. [3]. The natural measure (5.18) (more carefully defined than in the above sketch) is often referred to as the SBR or Sinai-Bowen-Ruelle measure [14, 13, 16]. The Heisenberg picture in dynamical system theory has been introduced in refs. [4, 5], see also ref. [3].

Remark **5.2** Koopman operators. Inspired by the contemporary advances in quantum mechanics, Koopman [4] observed in 1931 that $\mathcal{K}^t$ is unitary on $L^2(\mu)$ Hilbert spaces. The Liouville/Koopman operator is the classical analogue of the quantum evolution operator — the kernel of $\mathcal{L}^t(y,x)$ introduced in (5.16) (see also sect. 6.2) is the analogue of the Green's function. The relation between the spectrum of the Koopman operator and classical ergodicity was formalized by von Neumann [5]. We shall not use

Hilbert spaces here and the operators that we shall study *will not* be unitary. For a discussion of the relation between the Perron-Frobenius operators and the Koopman operators for finite dimensional deterministic invertible flows, infinite dimensional contracting flows, and stochastic flows, see Lasota-Mackey [3] and Gaspard [4].

Remark **5.3** Bounded semigroup. For a discussion of bounded semigroups of page 107 see, for example, Marsden and Hughes [6].

Résumé

In a chaotic system, it is not possible to calculate accurately the long time trajectory of a given initial point. We study instead the evolution of the *measure*, or the *density* of representative points in phase space, acted upon by an evolution operator. Essentially this means trading in nonlinear dynamical equations on finite low-dimensional spaces $x = (x_1, x_2 \cdots x_d)$ for linear equations on infinite dimensional vector spaces of density functions $\rho(x)$.

Reformulated this way, classical dynamics takes on a distinctly quantum-mechanical flavor. Both in classical and quantum mechanics one has a choice of implementing dynamical evolution on densities ("Schrödinger picture", sect. 5.4) or on observables ("Heisenberg picture", sect. 6.2 and chapter 7): 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.

For long times the dynamics is described in terms of stationary measures, that is, fixed points of certain evolution operators. The most physical of stationary measures is the natural measure, a measure robust under perturbations by weak noise.

References

[5.1] Ya.G. Sinai, *Topics in Ergodic Theory*, (Princeton University Press, Princeton, New Jersey, 1994).

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

[5.3] A. Lasota and M.C. Mackey, *Chaos, Fractals and Noise* (Springer, New York 1994).

[5.4] B.O. Koopman, *Proc. Nat. Acad. Sci. USA* **17**, 315 (1931).

[5.5] J. von Neumann, *Ann. Math.* **33**, 587 (1932).

[5.6] J.E. Marsden and T.J.R. Hughes, *Mathematical Foundations of Elasticity* (Prentice-Hall, Englewood Cliffs, New Jersey, 1983)

[5.7] P. Cvitanović, C.P. Dettmann, R. Mainieri and G. Vattay, *Trace formulas for stochastic evolution operators: Weak noise perturbation theory*, *J. Stat. Phys.* **93**, 981 (1998); `chao-dyn/9807034`.

Exercises

5.1 Integrating over Dirac delta functions. Let us verify a few of the properties of the delta function and check (5.9), as well as the formulas (5.7) and (5.8) to be used later.

(a) If $f : \mathbb{R}^d \to \mathbb{R}^d$, then show that

$$\int_{\mathbb{R}^d} dx\, \delta\left(f(x)\right) = \sum_{x \in f^{-1}(0)} \frac{1}{|\det \partial_x f|}\,.$$

(b) The delta function can be approximated by delta sequences, for example

$$\int dx\, \delta(x) f(x) = \lim_{\sigma \to 0} \int dx\, \frac{e^{-\frac{x^2}{2\sigma}}}{\sqrt{2\pi\sigma}} f(x)\,.$$

Use this approximation to see whether the formal expression

$$\int_{\mathbb{R}} dx\, \delta(x^2)$$

makes sense.

5.2 Derivatives of Dirac delta functions. Consider $\delta^{(k)}(x) = \frac{\partial^k}{\partial x^k}\delta(x)$, and show that

(a) Using integration by parts, determine the value of

$$\int_{\mathbb{R}} dx\, \delta'(y)\,.$$

where $y = f(x) - x$.

(b)
$$\int dx\, \delta^{(2)}\left(y\right) = \sum_{x:y(x)=0} \frac{1}{|y'|} \left\{ 3\frac{(y'')^2}{(y')^4} - \frac{y'''}{(y')^3} \right\}\,. \tag{5.38}$$

(c)
$$\int dx\, b(x)\delta^{(2)}(y) = \sum_{x:y(x)=0} \frac{1}{|y'|} \left\{ \frac{b''}{(y')^2} - \frac{b'y''}{(y')^3} + b\left(3\frac{(y'')^2}{(y')^4} - \frac{y'''}{(y')^3} \right) \right\}\,. \tag{5.39}$$

These formulas are useful incomputing effects of weak noise on deterministic dynamics [7].

5.3 **$\mathcal{L}^t$ generates a semigroup.** Check that the Perron-Frobenius operator has the semigroup property,

$$\int_M dz \mathcal{L}^{t_2}(y,z)\,\mathcal{L}^{t_1}(z,x) = \mathcal{L}^{t_2+t_1}(y,x)\,, \qquad t_1, t_2 \geq 0\,. \tag{5.40}$$

As the flows that we tend to be interested in are invertible, the $\mathcal{L}$'s that we will use often do form a group, with $t_1, t_2 \in \mathbb{R}$.

5.4 **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

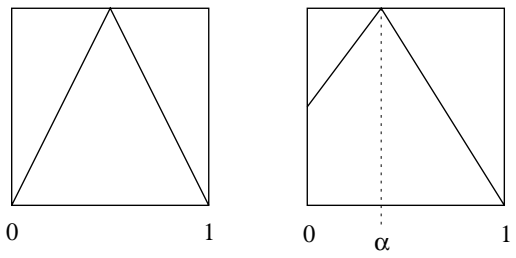
$$f(x) = a(1 - 2|x - 0.5|)$$

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).

5.5 **Invariant measure.** We will compute the invariant measure for two different piecewise linear maps.



(a) Verify the matrix $\mathcal{L}$ representation (5.13).

(b) The maximum of the first map has value 1. Compute an invariant measure for this map.

(c) Compute the leading eigenvalue of $\mathcal{L}$ for this map.

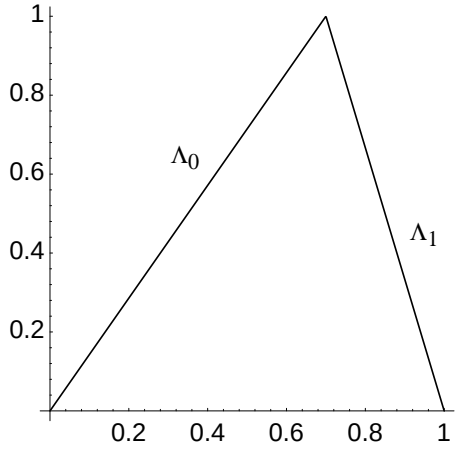
(d) For this map there is an infinite number of invariant measures, but only one of them will be found when one carries out a numerical simulation. Determine that measure, and explain why your choice is the natural measure for this map.

(e) In the second map the maximum is 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)

5.6 Escape rate for a flow conserving map. Adjust Λ_0, Λ_1 in (5.11) so that the gap between the intervals $\mathcal{M}_0$, $\mathcal{M}_1$ vanishes. Check that in that case the escape rate equals zero.

5.7 Eigenvalues of the skew Ulam tent map Perron-Frobenius operator. Show that for the skew Ulam tent map



$$f(x) = \begin{cases} f_0(x) = \Lambda_0 x\,, & x \in \mathcal{M}_0 = [0, 1/\Lambda_0) \\ f_1(x) = \frac{\Lambda_0}{\Lambda_0 - 1}(1 - x)\,, & x \in \mathcal{M}_1 = (1/\Lambda_0, 1]\,. \end{cases} \tag{5.41}$$

the eigenvalues are available analytically, compute the first few.

5.8 "Kissing disks"* (continuation of exercises 3.7 and 3.8). Close off the escape by setting $R = 2$, and look in the 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 around special regions of phase space for long times? These exemplify "intermittency", a bit of unpleasantness that we shall return to in chapter 16.

5.9 Invariant measure for the Gauss map. Consider the Gauss map (we shall need this map in chapter 19):

$$f(x) = \begin{cases} \frac{1}{x} - \left[\frac{1}{x}\right] & x \neq 0 \\ 0 & x = 0 \end{cases}$$

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?

5.10 Perron-Frobenius operator is the adjoint of the Koopman operator. Check (5.24) - it might be wrong as it stands. Pay attention to presence/absence of a Jacobian.

5.11 Exponential form of the semigroup. 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 phase space function $a(x)$.

5.12 $\mathcal{A}$ as a generator of translations. Verify that for a constant velocity field the evolution generator $\mathcal{A}$ n (5.28) is the generator of translations,

$$e^{tv\frac{\partial}{\partial x}}a(x) = a(x+tv)\,.$$

(hint: expand $a(x)$ in a Tylor series.)

5.13 Incompressible flows. Show that (5.9) 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 with the numerical experiment of exercise 5.8.

Chapter 6

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 start by discussing 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. 6.3 we try out the formalism on the first quantitative diagnosis that a system's got chaos, Lyapunove exponents.

6.1 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 phase 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.

6.1.1 Time averages

Let $a = a(x)$ be any *observable*, a function that associates to each point in phase 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 τ 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,

$$A^t(x_0) = \int_0^t d\tau\, a(f^\tau(x_0))\,. \tag{6.1}$$

If the dynamics is given by an iterated mapping and the time is discrete, $t \to n$, the integrated observable is given by

$$A^n(x_0) = \sum_{k=0}^{n-1} a(f^k(x_0)) \tag{6.2}$$

(we suppress possible vectorial indices for the time being). For example, if the observable is the velocity, $a_i(x) = v_i(x)$, its time integral $A_i^t(x_0)$ is the trajectory $A_i^t(x_0) = x_i(t)$. Another familiar example, for Hamiltonian flows, is the action associated with a trajectory $x(t) = [p(t), q(t)]$ passing through a phase space point $x_0 = [p(0), q(0)]$ (this function will be the key to the semiclassical quantization of chapter 22):

$$A^t(x_0) = \int_0^t d\tau\, \dot{\mathbf{q}}(\tau) \cdot \mathbf{p}(\tau)\,. \tag{6.3}$$

The *time average* of the observable along a trajectory is defined by

$$\overline{a(x_0)} = \lim_{t\to\infty} \frac{1}{t} A^t(x_0)\,. \tag{6.4}$$

If a does not behave too wildly as a function of time – for example, if $a_i(x)$ is the Chicago temperature, bounded between $-80^o F$ and $+130^o F$ for all times –

$A^t(x_0)$ is expected to grow not faster than t, and the limit (6.4) exists. For an example of a time average - the Lyapunov exponent - see sect. 6.3.

The time average depends on the trajectory, but not on the initial point on that trajectory: if we start at a later phase space point $f^T(x_0)$ we get a couple of extra finite contributions that vanish in the $t \to \infty$ limit:

$$\begin{aligned}
\overline{a(f^T(x_0))} &= \lim_{t\to\infty} \frac{1}{t} \int_T^{t+T} d\tau\, a(f^\tau(x_0)) \\
&= \overline{a(x_0)} - \lim_{t\to\infty} \frac{1}{t} \left(\int_0^T d\tau\, a(f^\tau(x_0)) - \int_t^{t+T} d\tau\, a(f^\tau(x_0)) \right) \\
&= \overline{a(x_0)}\,.
\end{aligned}$$

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

6.1 on p. 132

$$\begin{aligned}
\text{flows:} \quad A_p &= a_p T_p = \int_0^{T_p} a\left(f^\tau(x_0)\right) d\tau\,, \qquad x_0 \in p \\
\text{maps:} \quad &= a_p n_p = \sum_{i=0}^{n_p-1} a\left(f^i(x_0)\right)\,,
\end{aligned} \tag{6.5}$$

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 parcourse 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 (6.4) requires therefore only a single traversal of the cycle:

$$a_p = \frac{1}{T_p} A_p\,. \tag{6.6}$$

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 (6.6) on any periodic orbit, that is, on a dense set of points (fig. 6.1(b)). For example, for an open system such as the Sinai gas of sect. 18.1 (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 $x(t)^2 \sim$

chapter 18

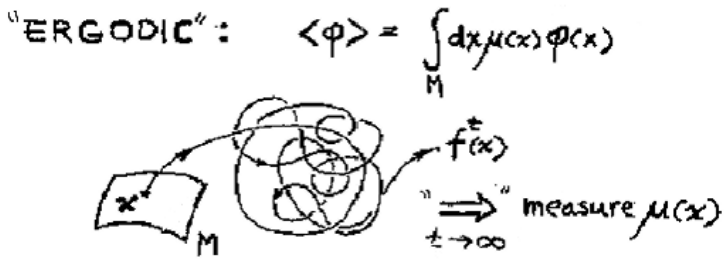


Figure 6.1: (a) A typical chaotic trajectory explores the phase space with the long time visitation frequency corresponding to the natural measure. (b) time average evaluated along an atypical trajectory such as a periodic orbit fails to explore the entire accessible phase space. (PC: clip out "Ergodic"; need to draw (b) here!)

t^2, and its contribution to the diffusion rate $D \approx x(t)^2/t$, (6.4) 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.

For chaotic dynamical systems, this paradox is resolved by robust averaging, that is, averaging also over the initial x, and worrying about the measure of the "pathological" trajectories.

6.1.2 Space averages

The *space average* of a quantity a that may depend on the point x of phase space $\mathcal{M}$ and on the time t is given by the d-dimensional integral over the d coordinates of the dynamical system:

$$\begin{aligned} \langle a \rangle(t) &= \frac{1}{|\mathcal{M}|}\int_{\mathcal{M}} dx\, a(x(t)) \\ |\mathcal{M}| &= \int_{\mathcal{M}} dx \;= \text{volume of } \mathcal{M}\,. \end{aligned} \tag{6.7}$$

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. 6.1.3).

What is it we *really* do in experiments? We cannot measure the time average (6.4), 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 (6.7), and consider instead

$$\langle a \rangle_\rho(t) = \frac{1}{|\mathcal{M}|}\int_{\mathcal{M}} dx\, \rho(x) a(x(t))\,. \tag{6.8}$$

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 (6.7); 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 phase space $\mathcal{M}$

$$\langle a \rangle = \lim_{t\to\infty} \frac{1}{|\mathcal{M}|}\int_{\mathcal{M}} dx\, \frac{1}{t}\int_0^t d\tau\, a(f^\tau(x))\,. \tag{6.9}$$

We use the same $\langle \cdots \rangle$ notation as for the space average (6.7), 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

$$\left\langle e^{\beta\cdot A^t} \right\rangle = \frac{1}{|\mathcal{M}|}\int_{\mathcal{M}} dx\, e^{\beta\cdot A^t(x)}. \tag{6.10}$$

In the present context β is an auxiliary variable of no particular physical significance. In most applications β is a scalar, but if the observable is a d-dimensional vector $a_i(x) \in \mathbb{R}^d$, then β is a conjugate vector $\beta \in \mathbb{R}^d$; if the observable is a $d \times d$ tensor, β is also a rank-2 tensor, and so on. Here we will mostly limit the considerations to scalar values of β.

If the limit $\overline{a(x_0)}$ for the time average (6.4) 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 (6.10) 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

$$s(\beta) = \lim_{t\to\infty} \frac{1}{t} \ln \left\langle e^{\beta \cdot A^t} \right\rangle . \tag{6.11}$$

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 (6.9) and the moments of the integrated observable (6.1) can be computed by evaluating the derivatives of $s(\beta)$

$$\begin{aligned} \left.\frac{\partial s}{\partial \beta}\right|_{\beta=0} &= \lim_{t\to\infty} \frac{1}{t} \langle A^t \rangle = \langle a \rangle , \\ \left.\frac{\partial^2 s}{\partial \beta^2}\right|_{\beta=0} &= \lim_{t\to\infty} \frac{1}{t} \left(\langle A^t A^t \rangle - \langle A^t \rangle \langle A^t \rangle \right) \\ &= \lim_{t\to\infty} \frac{1}{t} \left\langle (A^t - t\langle a \rangle)^2 \right\rangle , \end{aligned} \tag{6.12}$$

6.3 on p. 133 and so forth. We have written out the formulas for a scalar observable; the vector case is worked out in the exercise 6.3. 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.

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 $\langle x(t)^2 \rangle \approx 4Dt$. 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 $\langle v_i \rangle = 0$, the diffusion constant is given by the curvature of $s(\beta)$ at $\beta = 0$,

$$D = \lim_{t\to\infty} \frac{1}{2dt} \langle x(t)^2 \rangle = \frac{1}{2d} \sum_{i=1}^{d} \left.\frac{\partial^2 s}{\partial \beta_i^2}\right|_{\beta=0} , \tag{6.13}$$

sect. 18.1

so if we can evaluate derivatives of $s(\beta)$, we can compute transport coefficients that characterize deterministic diffusion. As we shall see in chapter 18, periodic orbit theory yields an explicit closed form expression for D.

fast track:
sect. 6.2, p. 124

6.1.3 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, (6.9) 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

$$\int_{\mathcal{M}} dy\, \delta(y - f^t(x_0)) = 1\,, \quad t > 0\,, \qquad \text{iff } x_0 \in \text{ non–wandering set} \tag{6.14}$$

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 phase 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 phase space containing all trajectories which start out within the phase space region $\mathcal{M}$ and recur within that region at the time t

$$|\mathcal{M}(t)| = \int_{\mathcal{M}} dxdy\, \delta\big(y - f^t(x)\big) \;\sim\; |\mathcal{M}|e^{-\gamma t} \tag{6.15}$$

is expected to decrease exponentially, with the escape rate γ. The integral over x takes care of all possible initial points; the integral over y checks whether their trajectories are still within $\mathcal{M}$ by the time t. For example, any trajectory that falls off the pinball table in fig. 1.1 is gone for good.

sect. 1.3.5
sect. 14.1

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 (6.15), by redefining the space average (6.10) as

$$\left\langle e^{\beta\cdot A^t} \right\rangle = \int_{\mathcal{M}} dx\, \frac{1}{|\mathcal{M}(t)|} e^{\beta\cdot A^t(x)} \sim \frac{1}{|\mathcal{M}|}\int_{\mathcal{M}} dx\, e^{\beta\cdot A^t(x)+\gamma t}\,. \tag{6.16}$$

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 γ 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, $\langle 1 \rangle = 1$.

We now turn to the problem of evaluating $\left\langle e^{\beta \cdot A^t} \right\rangle$.

6.2 Evolution operators

The above simple shift of focus, from studying $\langle a \rangle$ to studying $\left\langle \exp\left(\beta \cdot A^t\right) \right\rangle$ is the key to all that follows. Make the dependence on the flow explicit by rewriting this quantity as

$$\left\langle e^{\beta \cdot A^t} \right\rangle = \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} dx \int_{\mathcal{M}} dy \, \delta\big(y - f^t(x)\big) \, e^{\beta \cdot A^t(x)} \, . \tag{6.17}$$

Here $\delta\big(y - f^t(x)\big)$ 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

$$1 = \int_{\mathcal{M}} dy \, \delta\big(y - f^t(x)\big) \, , \tag{6.18}$$

into (6.10) 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 phase 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 (5.25) 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 $\mathcal{L}^t = e^{t\mathcal{A}}$ in the phase-space representation (6.17) as the *evolution operator*

$$\mathcal{L}^t(y,x) \, = \, \delta\big(y - f^t(x)\big) \, e^{\beta \cdot A^t(x)} \, . \tag{6.19}$$

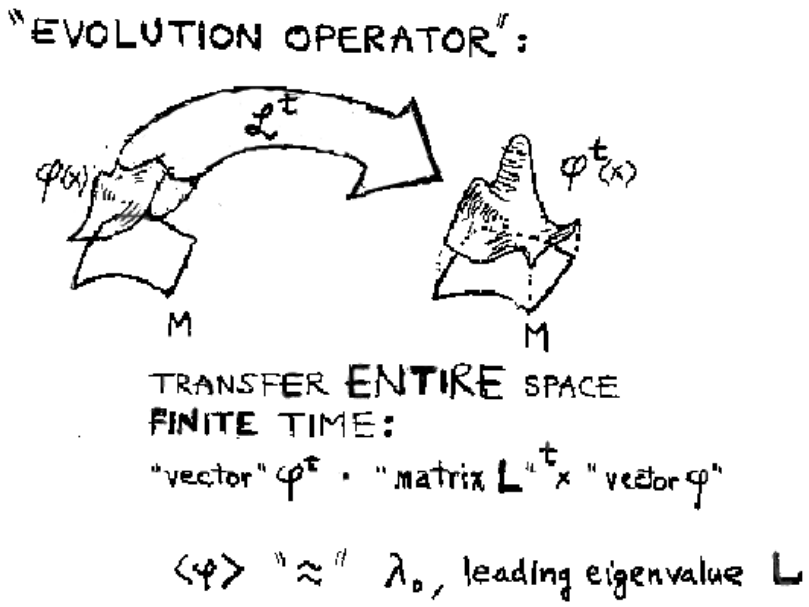


Figure 6.2: Space averaging pieces together the time average computed along the $t \to \infty$ trajectory of fig. 6.1 by a simultaneous space average over finite t trajectory segments starting at infinitely many starting points.

The simplest example is the Perron-Frobenius operator introduced in section 5.2. Another example - designed to deliver the Lyapunov exponent - will be the evolution operator (6.31). The evolution operator acts on scalar functions $a(x)$ as

$$\mathcal{L}^t a(y) = \int_{\mathcal{M}} dx\, \delta\big(y - f^t(x)\big)\, e^{\beta\cdot A^t(x)} a(x)\,. \tag{6.20}$$

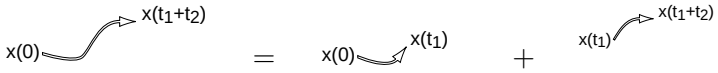
In terms of the evolution operator, the expectation value of the generating function (6.17) is given by

$$\left\langle e^{\beta\cdot A^t} \right\rangle = \left\langle \mathcal{L}^t \iota \right\rangle\,,$$

where the initial density $\iota(x)$ is the constant function that always returns 1.

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



$$\begin{aligned} A^{t_1+t_2}(x_0) &= \int_0^{t_1} d\tau\, a(x(\tau)) + \int_{t_1}^{t_1+t_2} d\tau\, a(x(\tau)) \\ &= A^{t_1}(x_0) + A^{t_2}(f^{t_1}(x_0))\,. \end{aligned}$$

6.2 on p. 132

If $A^t(x)$ is additive along the trajectory, the evolution operator generates a semigroup

sect. 5.4

$$\mathcal{L}^{t_1+t_2}(y,x) = \int_{\mathcal{M}} dz\, \mathcal{L}^{t_2}(y,z)\mathcal{L}^{t_1}(z,x)\,, \tag{6.21}$$

as is easily checked by substitution

$$\mathcal{L}^{t_2}\mathcal{L}^{t_1}a(x) = \int_{\mathcal{M}} dy\, \delta(x - f^{t_2}(y))e^{\beta\cdot A^{t_2}(y)}(\mathcal{L}^{t_1}a)(y) = \mathcal{L}^{t_1+t_2}a(x)\,.$$

This semigroup property is the main reason why (6.17) is preferable to (6.9) as a starting point for evaluation of dynamical averages: it recasts averaging in form of operators multiplicative along the flow.

6.3 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, two trajectories

sect. 1.3.1

$$x(t) = f^t(x_0) \quad \text{and} \quad x(t) + \delta x(t) = f^t(x_0 + \delta x(0)) \tag{6.22}$$

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 phase space. This *sensitivity to initial conditions* can be quantified as

$$|\delta x(t)| \approx e^{\lambda t}|\delta x(0)| \tag{6.23}$$

where λ, the mean rate of separation of trajectories of the system, is called the *Lyapunov exponent*.

6.3.1 Lyapunov exponent as a time average

We can start out with a small δx and try to estimate λ from (6.23), but now that we have quantified the notion of linear stability in chapter 4 and defined the dynamical time averages in sect. 6.1.1, 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 tangent vectors to a given orbit.

The mean growth rate of the distance $|\delta x(t)|/|\delta x(0)|$ between neighboring trajectories (6.22) is given by the *Lyapunov exponent*

$$\lambda = \lim_{t\to\infty} \frac{1}{t} \ln |\delta x(t)|/|\delta x(0)| \tag{6.24}$$

(For notational brevity we shall often suppress the dependence of $\lambda = \lambda(x_0)$ and related quantities on the initial point x_0 and the time t). For infinitesimal δx we know the $\delta x_i(t)/\delta x_j(0)$ ratio exactly, as this is by definition the Jacobian matrix (4.25)

$$\lim_{\delta x\to 0} \frac{\delta x_i(t)}{\delta x_j(0)} = \frac{\partial x_i(t)}{\partial x_j(0)} = \mathbf{J}^t_{ij}(x_0)\,,$$

so the leading Lyapunov exponent can be computed from the linear approximation (4.24)

$$\lambda = \lim_{t\to\infty} \frac{1}{t} \ln \frac{\left|\mathbf{J}^t(x_0)\delta x(0)\right|}{|\delta x(0)|} = \lim_{t\to\infty} \frac{1}{2t} \ln \left|\hat{n}^T (\mathbf{J}^t)^T \mathbf{J}^t \hat{n}\right| \,. \tag{6.25}$$

In this formula the scale of the initial separation drops out, only its orientation given by the unit vector $\hat{n} = \delta x/|\delta x|$ matters. The eigenvalues of $\mathbf{J}$ are either real or come in complex conjugate pairs. As $\mathbf{J}$ is in general not symmetric and not diagonalizable, it is more convenient to work with the symmetric and diagonalizable matrix $\mathbf{M} = (\mathbf{J}^t)^T \mathbf{J}^t$, with real 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 $\mathbf{M}u_i = \Lambda_i u_i$ eigenbasis, we have

$$\hat{n}^T \mathbf{M} \hat{n} = \sum_{i=1}^{d} (\hat{n}\cdot u_i)^2 |\Lambda_i|^2 = (\hat{n}\cdot u_1)^2 e^{2\lambda_1 t} \left(1 + O(e^{-2(\lambda_1-\lambda_2)t})\right) \,, \tag{6.26}$$

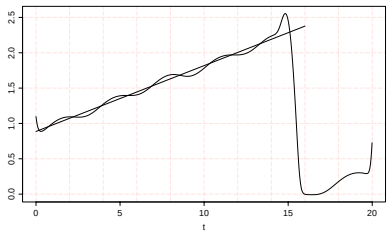


Figure 6.3: A numerical estimate of the leading Lyapunov exponent for the Rössler system (2.12) from the dominant expanding eigenvalue formula (6.25). The leading Lyapunov exponent $\lambda \approx 0.09$ is positive, so numerics supports the hypothesis that the Rössler attractor is strange. (J. Mathiesen)

where $t\lambda_i = \log|\Lambda_i(x_0,t)|$, and we assume that $\lambda_1 > \lambda_2 \geq \lambda_3 \cdots$. For long times the largest Lyapunov exponent dominates exponentially (6.25), 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

$$\begin{aligned} \overline{\lambda(x_0)} &= \lim_{t\to\infty} \frac{1}{t} \left\{ \log|\hat{n}\cdot u_1| + \log|\Lambda_1(x_0,t)| + O(e^{-2(\lambda_1-\lambda_2)t}) \right\} \\ &= \lim_{t\to\infty} \frac{1}{t} \log|\Lambda_1(x_0,t)|\,, \end{aligned} \tag{6.27}$$

where $\Lambda_1(x_0,t)$ is the leading eigenvalue of $\mathbf{J}^t(x_0)$. By chosing 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:

$$\overline{\lambda_i(x_0)} = \lim_{t\to\infty} \frac{1}{t} \ln|\Lambda_i(x_0,t)|\,, \qquad i = 1,2,\cdots,d\,. \tag{6.28}$$

The leading Lyapunov exponent now follows from the Jacobian matrix by numerical integration of (4.31). The equations can be integrated accurately for a finite time, hence the infinite time limit of (6.25) can be only estimated from plots of $\frac{1}{2}\ln|\hat{n}^T\mathbf{M}\hat{n}|$ as function of time, such as the fig. 6.3 for the Rössler system (2.12). 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 (6.27) directly. First of all, the phase space is dense with atypical trajectories; for example, if x_0 happened to lie on a periodic orbit p, $\overline{\lambda}$ would be simply $\log|\Lambda_p|/T_p$, a local property of cycle p, not a global property of the dynamical system. Furthermore, even if x_0 happens to be a "generic" phase space point, it is still not obvious that $\log|\Lambda(x_0,t)|/t$ 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 $\log|\Lambda(x_0,t)|/t$ can dip arbitrarily close to 0^+. For phase space volume non-preserving flows the trajectory can traverse locally contracting regions, and $\log|\Lambda(x_0,t)|/t$ 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.

4.1 on p. 94

6.3.2 Evolution operator evaluation of Lyapunov exponents

A cure to these problems was offered in sect. 6.2. We shall now replace time averaging along a single trajectory by action of a multiplicative evolution operator on the entire phase space, and extract the Lyapunov exponent from its leading eigenvalue. If the chaotic motion fills the whole phase 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 nonwandering 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 (4.52) for the derivative of an iterated map, the stability of a 1-d mapping is multiplicative along the flow, so the integral (6.1) of the observable $a(x) = \log|f'(x)|$, the local trajectory divergence rate, evaluated along the trajectory of x_0 is additive:

$$A^n(x_0) = \log\left|f^{n\prime}(x_0)\right| = \sum_{k=0}^{n-1} \log\left|f'(x_k)\right| \,. \tag{6.29}$$

The Lyapunov exponent is then the expectation value (6.9) given by a spatial integral (5.24) weighted by the natural measure

$$\lambda = \left\langle \log|f'(x)| \right\rangle = \int_{\mathcal{M}} dx\, \rho_0(x) \log|f'(x)| \,. \tag{6.30}$$

The associated (discrete time) evolution operator (6.19) is

$$\mathcal{L}(y,x) \;=\; \delta(y - f(x))\; e^{\beta \log|f'(x)|} \,. \tag{6.31}$$

appendix G.1

Here we have restricted our considerations to 1-dimensional maps, as for higher-dimensional flows only the Jacobian matrices are multiplicative, not the individual eigenvalues. Construction of the evolution operator for evaluation of the Lyapunov spectra 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

$$\lambda = \left\langle \log |f'(x)| \right\rangle = \left. \frac{\partial s(\beta)}{\partial \beta} \right|_{\beta=1} = s'(1) \tag{6.32}$$

from (6.12), the derivative of the leading eigenvalue $s_0(\beta)$ of the evolution operator (6.31). The only question is: how?

☞ sect. 13.2

in depth:
appendix G.1, p. 643

Commentary

Remark **6.1** "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. Either Ruelle [2] or Bowen [1] had given name "pressure" $P(a)$ (where a is the observable introduced here in sect. 6.1.1) to $s(\beta)$, defined by the "large system" limit (6.11). For us, $s(\beta)$ will be the leading eigenvalue of the evolution operator introduced in sect. 5.4, and the "convexity" properties such as $P(a) \leq P(|a|)$ will be pretty obvious consequence of the definition (6.11). In physics vernacular the eigenvalues $\{s_0(\beta), s_1(\beta), \cdots\}$ in the case that $\mathcal{L}$ is the Perron-Frobenius operator (5.10) are called the Ruelle-Pollicott resonances, 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, Ω-sets which are not really essential to understanding the physical applications of the theory, and are avoided in this presentation.

Remark **6.2** Microcanonical ensemble. In statistical mechanics the space average (6.7) performed over the Hamiltonian system constant energy surface invariant measure $\rho(x)dx = dqdp\,\delta(H(q,p) - E)$ of volume $|\mathcal{M}| = \int_{\mathcal{M}} dqdp\,\delta(H(q,p) - E)$

$$\langle a(t) \rangle = \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} dqdp\,\delta(H(q,p) - E) a(q,p,t) \tag{6.33}$$

is called the *microcanonical ensemble average*.

Remark **6.3** Lyapunov exponents. The Multiplicative Ergodic Theorem of Oseledec states that the limit (6.28) exists for almost all points x_0 and all tangent vectors $\hat{n}$. There are at most d distinct values of λ as we let $\hat{n}$ range over the tangent space. These are the Lyapunov exponents $\lambda_i(x_0)$.

There is a rather large literature on numerical computation of the Lyapunov exponents, see for example refs. [3, 4].

Résumé

The expectation value $\langle a \rangle$ of an observable $a(x)$ measured and averaged along the flow $x \to f^t(x)$ is given by the derivative $\partial s/\partial\beta$ of the leading eigenvalue $e^{ts(\beta)}$ of the evolution operator $\mathcal{L}^t$.

Next question is: how do we evalute the eigenvalues of $\mathcal{L}$? We saw in sect. 5.2.1, 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. As we shall show in chapters 7 and 8, a systematic way to accomplish this task is by means of periodic orbits.

References

[6.1] R.Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Springer Lecture Notes on Mathematics **470** (1975)

[6.2] D. Ruelle, *"Statistical mechanics of a one-dimensional lattice gas"*, *Commun. Math. Phys.* **9**, 267 (1968).

[6.3] Wolf, A., J. B. Swift, et al. (1985). "Determining Lyapunov Exponents from a Time Series." Physica D 16: 285-317.

[6.4] Eckmann, J.-P., S. O. Kamphorst, et al. (1986). "Liapunov exponents from time series." Phys. Rev. A 34: 4971-4979.

Exercises

6.1 A contracting baker's map. Consider a contracting (or "dissipative") baker's map, on $[0,1]^2$, defined as

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

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

This map shrinks strips by factor $1/3$ in the x direction, and stretches (and folds) by factor 2 in the y direction.

(a) How fast does the phase space volume contract?

(b) The symbolic dynamics encoding of trajectories is realized via symbols 0 ($y \leq 1/2$) and 1 ($y > 1/2$). Consider the observable $a(x,y) = x$. Verify that for any periodic orbit p $(\epsilon_1 \dots \epsilon_{n_p})$, $\epsilon_i \in \{0,1\}$

$$A_p = \frac{3}{4} \sum_{j=1}^{n_p} \delta_{j,1} \,.$$

6.2 $\mathcal{L}^t$ generates a semigroup. Check that the evolution operator has the semigroup property,

$$\int_M dz \mathcal{L}^{t_2}(y,z)\, \mathcal{L}^{t_1}(z,x) = \mathcal{L}^{t_2+t_1}(y,x)\,, \qquad t_1, t_2 \geq 0\,. \tag{6.34}$$

As the flows that we tend to be interested in are invertible, the $\mathcal{L}$'s that we will use often do form a group, with $t_1, t_2 \in \mathbb{R}$.

6.3 Expectation value of a vector observable and its moments. Check and extend the expectation value formulas (6.12) by evaluating the derivatives of $s(\beta)$ up to 4-th order for the space average $\left\langle \exp(\beta \cdot A^t) \right\rangle$ with a_i a vector quantity:

(a)

$$\left.\frac{\partial s}{\partial \beta_i}\right|_{\beta=0} = \lim_{t\to\infty}\frac{1}{t}\left\langle A_i^t\right\rangle = \langle a_i\rangle \ , \tag{6.35}$$

(b)

$$\begin{aligned}\left.\frac{\partial^2 s}{\partial\beta_i\partial\beta_j}\right|_{\beta=0} &= \lim_{t\to\infty}\frac{1}{t}\left(\left\langle A_i^t A_j^t\right\rangle - \left\langle A_i^t\right\rangle\left\langle A_j^t\right\rangle\right) \\ &= \lim_{t\to\infty}\frac{1}{t}\left\langle (A_i^t - t\langle a_i\rangle)(A_j^t - t\langle a_j\rangle)\right\rangle \ . \end{aligned}\tag{6.36}$$

Note that the formalism is cmart: 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 (6.35) vanishes, $\langle a_i\rangle = 0$. The 4-th order moment formula

$$K(t) = \frac{\left\langle x^4(t)\right\rangle}{\left\langle x^2(t)\right\rangle^2} - 3 \tag{6.37}$$

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 (18.21) for a concrete example). If the observable is a vector, the kurtosis is given by

$$K(t) = \frac{\sum_{ij}\left[\langle A_iA_iA_jA_j\rangle + 2\left(\langle A_iA_j\rangle\langle A_jA_i\rangle - \langle A_iA_i\rangle\langle A_jA_j\rangle\right)\right]}{\left(\sum_i\langle A_iA_i\rangle\right)^2} \tag{6.38}$$

Chapter 7

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. 1.4 and 1.5; now we redo the same material in greater depth. In chapter 6 we have related global averages to the eigenvalues of appropriate evolution operators. Traces of evolution operators can be evaluated as integrals over Dirac delta functions, and in this way the spectra of evolution operators become related to periodic orbits. If there is one idea that one should learn about chaotic dynamics, it happens in this chapter, and it is this: there is a fundamental local $\leftrightarrow$ global duality which says that

the spectrum of eigenvalues is dual to the spectrum of periodic orbits

For dynamics on the circle, this is called Fourier analysis; for dynamics on well-tiled manifolds, Selberg traces and zetas; and for generic nonlinear dynamical systems the duality is embodied in the trace formulas that we will now introduce. These objects are to dynamics what partition functions are to statistical mechanics.

7.1 Trace of an evolution operator

Our extraction of the spectrum of $\mathcal{L}$ commences with the evaluation of the trace. To compute an expectation value using (6.17) we have to integrate over all the values of the kernel $\mathcal{L}^t(x,y)$. If $\mathcal{L}^t$ were a matrix we would be computing a

weighted sum of its eigenvalues which is dominated by the leading eigenvalue as $t \to \infty$. As the trace of $\mathcal{L}^t$ is also dominated by the leading eigenvalue as $t \to \infty$, we might just as well look at the trace

11.2 on p. 260

$$\mathrm{tr}\, \mathcal{L}^t = \int dx\, \mathcal{L}^t(x,x) = \int dx\, \delta\big(x - f^t(x)\big)\, e^{\beta \cdot A^t(x)}\,. \tag{7.1}$$

Assume that $\mathcal{L}$ has a spectrum of discrete eigenvalues $s_0, s_1, s_2, \cdots$ ordered so that $\mathrm{Re}\, s_\alpha \geq \mathrm{Re}\, s_{\alpha+1}$. We ignore for the time being the question of what function space the eigenfunctions belong to, as we shall compute the eigenvalue spectrum without constructing any explicit eigenfunctions.

By definition, the trace is the sum over eigenvalues (for the time being we choose not to worry about convergence of such sums),

$$\mathrm{tr}\, \mathcal{L}^t = \sum_{\alpha=0}^{\infty} e^{s_\alpha t}\,. \tag{7.2}$$

On the other hand, we have learned in sect. 5.2 how to evaluate the delta-function integral (7.1).

As the case of discrete time mappings is somewhat simpler, we first derive the trace formula for maps, and then for flows. The final formula (7.19) covers both cases.

7.1.1 Hyperbolicity assumption

According to (5.8) the trace (7.1) picks up a contribution whenever $x - f^n(x) = 0$, that is whenever x belongs to a periodic orbit. For reasons which we will explain in sect. 7.1.4, 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,

$$\mathrm{tr}\, {}_p\mathcal{L}^{n_p} = \int_{\mathcal{M}_p} dx\, \delta(x - f^{n_p}(x)) = \frac{n_p}{\left|\det\, (\mathbf{1} - \mathbf{J}_p)\right|} = n_p \prod_{i=1}^{d} \frac{1}{|1 - \Lambda_{p,i}|} \tag{7.3}$$

(in (5.9) and here we set the observable $e^{A_p} = 1$ for the time being). Periodic orbit Jacobian matrix $\mathbf{J}_p$ is also known as the monodromy matrix (from Greek *mono-* = alone, single, and *dromo* = run, racecourse), and its eigenvalues $\Lambda_{p,1}$, $\Lambda_{p,2}$, ..., $\Lambda_{p,d}$ as the Floquet multipliers. We sort the eigenvalues $\Lambda_{p,1}$, $\Lambda_{p,2}$,

..., $\Lambda_{p,d}$ of the p-cycle $[d\times d]$ Jacobian matrix $\mathbf{J}_p$ into expanding, marginal and contracting sets $\{e, m, c\}$, as in (4.59). As the integral (7.3) can be carried out only if $\mathbf{J}_p$ has no eigenvalue of unit magnitude, we assume that no eigenvalue is marginal (we shall show in sect. 7.1.4, the longitudinal $\Lambda_{p,d+1} = 1$ eigenvalue for flows can be eliminated by restricting the consideration to the transverse Jacobian matrix $\mathbf{J}_p$), and factorize the trace (7.3) into a product over the expanding and the contracting eigenvalues

$$\left|\det\left(\mathbf{1}-\mathbf{J}_p\right)\right|^{-1} = \frac{1}{|\Lambda_p|}\prod_e \frac{1}{1-1/\Lambda_{p,e}}\prod_c \frac{1}{1-\Lambda_{p,c}}\ , \tag{7.4}$$

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:

$$\begin{aligned} |\Lambda_{p,e}| &> e^{\lambda_e T_p} \qquad \text{any } p\text{, any expanding eigenvalue } |\Lambda_{p,e}|>1 \\ |\Lambda_{p,c}| &< e^{-\lambda_c T_p} \qquad \text{any } p\text{, any contracting eigenvalue } |\Lambda_{p,c}|<1\,, \end{aligned} \tag{7.5}$$

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 16.

It follows from (7.4) that for long times, $t = rT_p \to \infty$, only the product of expanding eigenvalues matters, $\left|\det\left(\mathbf{1}-\mathbf{J}_p^r\right)\right| \to |\Lambda_p|^r$. We shall use this fact to motivate the construction of dynamical zeta functions in sect. 8.3. However, for evaluation of the full spectrum the exact cycle weight (7.3) has to be kept.

7.1.2 A 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:

$$\operatorname{tr}\mathcal{L}^n = \int dx\,\mathcal{L}^n(x,x) = \sum_{x_i\in \operatorname{Fix} f^n} \frac{e^{\beta\cdot A_i}}{\left|\det\left(\mathbf{1}-\mathbf{J}^n(x_i)\right)\right|}\,. \tag{7.6}$$

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 (6.5) 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 (5.8) by taking the determinant of $\partial_i(x_j - f^n(x)_j)$. If a trajectory retraces itself r times, its Jacobian matrix is $\mathbf{J}_p^r$, where $\mathbf{J}_p$ is the $[d\times d]$ Jacobian matrix (4.5) evaluated along a single traversal of the prime cycle p. As we saw in (6.5), 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. ☞ chapter ?? 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 (7.6) can be rewritten as a sum over all prime cycles and their repeats

$$\mathrm{tr}\,\mathcal{L}^n = \sum_p n_p \sum_{r=1}^{\infty} \frac{e^{r\beta\cdot A_p}}{\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right|}\delta_{n,n_p r}\,, \tag{7.7}$$

with the Kronecker delta $\delta_{n,n_p r}$ projecting out the periodic contributions of total period n. This constraint is awkward, and will be more awkward still for the continuous time flows, where it will yield a series of Dirac delta spikes (7.17). Such sums are familiar from the density-of-states sums of statistical mechanics, where they are dealt with in the same way as we shall do here: we smooth this distribution by taking a Laplace transform which rids us of the $\delta_{n,n_p r}$ constraint.

We *define* the trace formula for maps to be the Laplace transform of $\mathrm{tr}\,\mathcal{L}^n$ which, for discrete time mappings, is simply the generating function for the trace sums

$$\sum_{n=1}^{\infty} z^n \mathrm{tr}\,\mathcal{L}^n = \mathrm{tr}\,\frac{z\mathcal{L}}{1 - z\mathcal{L}} = \sum_p n_p \sum_{r=1}^{\infty} \frac{z^{n_p r} e^{r\beta\cdot A_p}}{\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right|}\,. \tag{7.8}$$

Expressing the trace as in (7.2), in terms of the sum of the eigenvalues of $\mathcal{L}$, we obtain the *trace formula for maps*:

$$\sum_{\alpha=0}^{\infty} \frac{ze^{s_\alpha}}{1 - ze^{s_\alpha}} = \sum_p n_p \sum_{r=1}^{\infty} \frac{z^{n_p r}\, e^{r\beta\cdot A_p}}{\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right|}\,. \tag{7.9}$$

This is our first example of the duality between the spectrum of eigenvalues and the spectrum of periodic orbits, announced in the introduction to this chapter.

fast track:
sect. 7.1.4, p. 140

7.1.3 A trace formula for transfer operators

For a piecewise-linear map (5.11), we can explicitely 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 (7.6) reduces to

$$\mathrm{tr}\,\mathcal{L}^n = \sum_{m=0}^{n} \binom{n}{m} \frac{1}{|1-\Lambda_0^m\Lambda_1^{n-m}|} = \sum_{k=0}^{\infty} \left(\frac{1}{|\Lambda_0|\Lambda_0^k} + \frac{1}{|\Lambda_1|\Lambda_1^k} \right)^n . \tag{7.10}$$

The eigenvalues are simply

$$e^{s_k} = \frac{1}{|\Lambda_0|\Lambda_0^k} + \frac{1}{|\Lambda_1|\Lambda_1^k} . \tag{7.11}$$

For $k = 0$ this is in agreement with the explicit transfer matrix (5.13) eigenvalues (5.14).

Alert reader should experience anxiety at this point. Is it not true that we have already written down explicitely the transfer operator in (5.13), 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 (5.13) is the correct operator on the space of functions piecewise constant on the two defining intervals $\{\mathcal{M}_0, \mathcal{M}_1\}$; on this space the operator indeed has only the eigenvalue e^{s_0}. As we shall see in sect. 9.1, the full spectrum (7.11) 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 (5.11) follows from (7.8)

$$\mathrm{tr}\,\frac{z\mathcal{L}}{1-z\mathcal{L}} = \frac{z\left(\frac{1}{|\Lambda_0-1|}+\frac{1}{|\Lambda_1-1|}\right)}{1-z\left(\frac{1}{|\Lambda_0-1|}+\frac{1}{|\Lambda_1-1|}\right)} , \tag{7.12}$$

verifying the trace formula (7.9).

7.1.4 A trace formula for flows

> Amazing! I did not understand a single word.
> Fritz Haake
>
> (R. Artuso and P. Cvitanović)

As any pair of nearby points on a cycle returns to itself exactly at each cycle period, the eigenvalue of the Jacobian matrix corresponding to the eigenvector along the flow necessarily equals unity for all periodic orbits. Hence for flows the trace integral $\mathrm{tr}\,\mathcal{L}^t$ requires a separate treatment for the longitudinal direction. To evaluate the contribution of an isolated prime cycle p of period T_p, restrict the integration to an infinitesimally thin tube $\mathcal{M}_p$ enveloping the cycle (see fig. 1.9), and choose a local coordinate system with a longitudinal coordinate $dx_\parallel$ along the direction of the flow, and d transverse coordinates $x_\perp$

$$\mathrm{tr}\,_p\mathcal{L}^t = \int_{\mathcal{M}_p} dx_\perp dx_\parallel\, \delta\big(x_\perp - f^t_\perp(x)\big)\, \delta\Big(x_\parallel - f^t_\parallel(x)\Big)\ . \tag{7.13}$$

(here we again set the observable $\exp(\beta\cdot A^t)=1$ for the time being). Let $v(x)$ be the magnitude of the velocity at the point x along the flow. $v(x)$ is strictly positive, as otherwise the orbit would stagnate for infinite time at $v(x)=0$ points, and that would get us nowhere. Therefore we can parametrize the longitudinal coordinate $x_\parallel$ by the flight time

$$x_\parallel(\tau) \;=\; \int_0^\tau d\sigma\, v(\sigma)\Big|_{mod\, L_p}$$

where $v(\sigma) = v(x_\parallel(\sigma))$, and L_p is the length of the circuit on which the periodic orbit lies (for the time being the *mod* operation in the above definition is redundant, as $\tau\in[0,T_p]$). With this parametrization

$$\Big(f^t_\parallel(x) - x_\parallel\Big) = \int_\tau^{t+\tau} d\sigma\, v(\sigma)\Big|_{mod\, L_p}$$

so that the integral around the longitudinal coordinate is rewritten as

$$\int_0^{L_p} dx_\parallel\, \delta\Big(x_\parallel - f^t_\parallel(x)\Big) \;=\; \int_0^{T_p} d\tau\, v(\tau)\, \delta\left(\int_\tau^{t+\tau} d\sigma\, v(\sigma)\Big|_{mod\, L_p}\right). \tag{7.14}$$

Now we notice that the zeroes of the argument of the delta function do not depend on τ, as v is positive, so we may rewrite (7.14) as

$$\int_0^{L_p} dx_\parallel \, \delta\Big(x_\parallel - f_\parallel^t(x)\Big) \;=\; \sum_{r=1}^{\infty} \delta(t - rT_p) \int_0^{T_p} d\tau \, v(\tau) \frac{1}{v(\tau + t)},$$

having used (5.7). The r sum starts from one as we are considering strictly positive times. Now we use another elementary property of delta functions, namely that

$$h(x)\delta(x - x_0) \;=\; h(x_0)\delta(x - x_0)$$

so that velocities cancel, and we get

$$\oint_p dx_\parallel \, \delta\Big(x_\parallel - f_\parallel^t(x)\Big) \;=\; T_p \sum_{r=1}^{\infty} \delta(t - rT_p) \;. \tag{7.15}$$

The fact that it is the prime period which arises also for repeated orbits comes from the fact that the space integration just sweeps once the circuit in phase space: a similar observation will be important for the derivation of the semiclassical trace formula in chapter 22. For the remaining transverse integration variables the Jacobian is defined in a reduced Poincaré surface of section $\mathcal{P}$ of constant $x_\parallel$. Linearization of the periodic flow transverse to the orbit yields

$$\int_{\mathcal{P}} dx_\perp \delta\Big(x_\perp - f_\perp^{rT_p}(x)\Big) = \frac{1}{\left|\det\, \left(\mathbf{1} - \mathbf{J}_p^r\right)\right|} \;, \tag{7.16}$$

where $\mathbf{J}_p$ is the p-cycle $[d\times d]$ *transverse* Jacobian matrix, and as in (7.5) we have to assume hyperbolicity, that is that the magnitudes of all transverse eigenvalues are bounded away from unity.

Substituting (7.15), (7.16) into (7.13), we obtain an expression for $\mathrm{tr}\, \mathcal{L}^t$ as a sum over all prime cycles p and their repetitions

$$\mathrm{tr}\, \mathcal{L}^t = \sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r\beta\cdot A_p}}{\left|\det\, \left(\mathbf{1} - \mathbf{J}_p^r\right)\right|} \delta(t - rT_p) \;. \tag{7.17}$$

A trace formula follows by taking a Laplace transform. This is a delicate step, since the transfer 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 ϵ smaller than the period of any periodic orbit, and write

$$\begin{aligned}\int_\epsilon^\infty dt\, e^{-st}\,\mathrm{tr}\,\mathcal{L}^t &= \mathrm{tr}\,\frac{e^{-(s-\mathcal{A})\epsilon}}{s-\mathcal{A}} = \sum_{\alpha=0}^{\infty}\frac{e^{-(s-s_\alpha)\epsilon}}{s-s_\alpha} \\ &= \sum_p T_p \sum_{r=1}^{\infty}\frac{e^{r(\beta\cdot A_p - sT_p)}}{\left|\det\left(\mathbf{1}-\mathbf{J}_p^r\right)\right|}\,,\end{aligned} \tag{7.18}$$

where $\mathcal{A}$ is the generator of the semigroup of dynamical evolution, sect. 5.4. The *classical trace formula for flows* is the $\epsilon\to\infty$ limit of the above expression:

$$\sum_{\alpha=0}^{\infty}\frac{1}{s-s_\alpha} = \sum_p T_p \sum_{r=1}^{\infty}\frac{e^{r(\beta\cdot A_p - sT_p)}}{\left|\det\left(\mathbf{1}-\mathbf{J}_p^r\right)\right|}\,. \tag{7.19}$$

7.1 on p. 146

This is 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. We see that the trace formula for maps (7.9) is a special case of the trace formula for flows. The relation between the continuous and discrete time cases can be summarized as follows:

$$\begin{aligned} T_p &\leftrightarrow n_p \\ e^{-s} &\leftrightarrow z \\ e^{t\mathcal{A}} &\leftrightarrow \mathcal{L}^n\,. \end{aligned} \tag{7.20}$$

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 (7.19) by methods such as Padé approximants. However, it pays to first perform a simple resummation which converts this divergence of a trace into a *zero* of a spectral determinant. We shall do this in sect. 8.2, after we complete our offering of trace formulas.

7.2 An asymptotic trace formula

In order to illuminate the manipulations of sect. 7.1.2 and relate them to something we already possess intuition about, we now rederive the heuristic sum of sect. 1.4.1 from the exact trace formula (7.9). The Laplace transforms (7.9) or (7.19) are designed to capture the time $\to\infty$ asymptotic behavior of the trace

sums. By the hyperbolicity assumption (7.5) for $t = T_p r$ large the cycle weight approaches

$$\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right| \to |\Lambda_p|^r\,, \tag{7.21}$$

where Λ_p is the product of the expanding eigenvalues of $\mathbf{J}_p$. Denote the corresponding approximation to the nth trace (7.6) by

$$\Gamma_n = \sum_i^{(n)} \frac{1}{|\Lambda_i|}\,, \tag{7.22}$$

and denote the approximate trace formula obtained by replacing the cycle weights $\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right|$ by $|\Lambda_p|^r$ in (7.9) by $\Gamma(z)$. Equivalently, think of this as a replacement of the evolution operator (6.19) by a transfer operator (as in sect. 7.1.3). For concreteness consider a dynamical system whose symbolic dynamics is complete binary, for example the 3-disk system fig. 1.3. In this case distinct periodic points that contribute to the nth periodic points sum (7.7) are labelled by all admissible itineraries composed of sequences of letters $s_i \in \{0, 1\}$:

$$\begin{aligned}
\Gamma(z) &= \sum_{n=1}^{\infty} z^n \Gamma_n = \sum_{n=1}^{\infty} z^n \sum_{x_i \in \mathrm{Fix} f^n} \frac{e^{\beta\cdot A^n(x_i)}}{|\Lambda_i|} \\
&= z\left\{\frac{e^{\beta\cdot A_0}}{|\Lambda_0|} + \frac{e^{\beta\cdot A_1}}{|\Lambda_1|}\right\} + z^2\left\{\frac{e^{2\beta\cdot A_0}}{|\Lambda_0|^2} + \frac{e^{\beta\cdot A_{01}}}{|\Lambda_{01}|} + \frac{e^{\beta\cdot A_{10}}}{|\Lambda_{10}|} + \frac{e^{2\beta\cdot A_1}}{|\Lambda_1|^2}\right\} \\
&\quad + z^3\left\{\frac{e^{3\beta\cdot A_0}}{|\Lambda_0|^3} + \frac{e^{\beta\cdot A_{001}}}{|\Lambda_{001}|} + \frac{e^{\beta\cdot A_{010}}}{|\Lambda_{010}|} + \frac{e^{\beta\cdot A_{100}}}{|\Lambda_{100}|} + \ldots\right\}
\end{aligned} \tag{7.23}$$

Both the cycle averages A_i and the stabilities Λ_i are the same for all points $x_i \in p$ in a cycle p. Summing over repeats of all prime cycles we obtain

$$\Gamma(z) = \sum_p \frac{n_p t_p}{1 - t_p}\,, \qquad t_p = z^{n_p} e^{\beta\cdot A_p}/|\Lambda_p|\,. \tag{7.24}$$

This is precisely our initial heuristic estimate (1.8). Note that we could not perform such sum over r in the exact trace formula (7.9) as $\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right| \neq \left|\det\left(\mathbf{1} - \mathbf{J}_p\right)\right|^r$; 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 (8.9). ☞ sect. 8.2

If the weights $e^{\beta A^n(x)}$ are multiplicative along the flow, and the flow is hyperbolic, for given β 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 nth level sum (7.6) tends to the leading $\mathcal{L}^n$ eigenvalue e^{ns_0}. Summing this asymptotic estimate level by level

$$\Gamma(z) \approx \sum_{n=1}^{\infty} (ze^{s_0})^n = \frac{ze^{s_0}}{1 - ze^{s_0}} \tag{7.25}$$

we see that we should be able to determine s_0 by determining the smallest value of $z = e^{-s_0}$ for which the cycle expansion (7.24) diverges.

If one is interested only in the leading eigenvalue of $\mathcal{L}$, it suffices to consider the approximate trace $\Gamma(z)$. We will use this fact below to motivate the introduction of dynamical zeta functions (8.11), and in sect. 8.5.1 we shall give the exact relation between the exact and the approximate trace formulas.

Commentary

Remark **7.1** Who's dunne it? Continuous time flow traces weighted by the cycle periods were introduced by Bowen [1] who treated them as Poincaré section suspensions weighted by the "time ceiling" function (3.2). They were used by Parry and Pollicott [2]. The derivation presented here [3] was designed to parallel as closely as possible the derivation of the Gutzwiller semiclassical trace formula, chapters **??** and 22.

Remark **7.2** 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 (7.1) 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 (7.6) are called *flat* traces (see the review **??** and chapter 9). Other names for traces appear as well: for instance, in the context of $1-d$ mappings, *sharp* traces refer to generalizations of (7.6) where contributions of periodic points are weighted by the Lefschetz sign ± 1, reflecting whether the periodic point sits on a branch of nth iterate of the map which crosses the diagonal starting from below or starting from above [12]. Such traces are connected to the theory of kneading invariants (see ref. [4] and references therein). Traces weighted by ± 1 sign of the derivative of the fixed point have been used to study the period doubling repeller, leading to high precision estimates of the Feigenbaum constant δ, refs. [5, 5, 6].

Résumé

The description of a chaotic dynamical system in terms of cycles can be visualized as a tessellation of the dynamical system, fig. 1.8, 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 phase space yields the *classical trace formula*

$$\sum_{\alpha=0}^{\infty} \frac{1}{s-s_\alpha} = \sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r(\beta\cdot A_p - sT_p)}}{\left|\det\left(\mathbf{1}-\mathbf{J}_p^r\right)\right|}\,.$$

Now that we have a trace formula we might ask what it is good for? It's not good for much as it stands, a scary 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 starting to grasp that the theory might turn out to be convergent beyond our wildest dreams (chapter 9). In order to implement such formulas, we have to determine "all" prime cycles. This task we postpone to chapters **??** and 12.

References

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

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

[7.3] P. Cvitanović and B. Eckhardt, *J. Phys.* **A 24**, L237 (1991).

[7.4] V. Baladi and D. Ruelle, *Ergodic Theory Dynamical Systems* **14**, 621 (1994).

[7.5] R. Artuso, E. Aurell and P. Cvitanović, *Nonlinearity* **3**, 325 (1990); *ibidem* 361 (1990)

[7.6] M. Pollicott, *J. Stat. Phys.* **62**, 257 (1991).

Exercises

7.1 $t \to 0_+$ **regularization of eigenvalue sums******.** In taking the Laplace transform (7.19) 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 (see sect. 22.1.1). Regularize the divergent sum in (7.19) following (for example) the prescription of appendix J.5 and assign to such volume term some interesting role in the theory of classical resonance spectra. E-mail the solution to the authors.

7.2 **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 $\mathbb{R}^{n_1} \mapsto \mathbb{R}^{n_2}$? ($n_i \in \mathbb{N}$.)

Chapter 8

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 trace formulas (7.9), (7.19) and (7.24) is that they diverge at $z = e^{-s_0}$, respectively $s = s_0$, that is, 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 of chapter 22 this leads to a perplexing observation that crude estimates of the radius of convergence seem to put the entire physical spectrum out of reach (see chapter 9). We shall now cure this problem by going from trace formulas to determinants. The idea is illustrated by fig. 1.10: Determinants tend to have larger analyticity domains because if $\operatorname{tr} \mathcal{L}/(1 - z\mathcal{L}) = \frac{d}{dz} \ln \det (1 - z\mathcal{L})$ 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 than its radius of convergence.

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 $\log \det = \operatorname{tr} \log$. 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.

8.1 Spectral determinants for maps

The eigenvalues z_k of a linear operator are given by the zeros of the determinant

$$\det(1 - z\mathcal{L}) = \prod_k (1 - z/z_k)\,. \tag{8.1}$$

For finite matrices this is the characteristic determinant; for operators this is the Hadamard representation of the *spectral determinant* (here again we spare 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

1.3 on p. 32

$$\ln\det(1 - M) = \mathrm{tr}\,\ln(1 - M) = -\sum_{n=1}^{\infty} \frac{1}{n}\mathrm{tr}\,M^n\,, \tag{8.2}$$

to relate the spectral determinant of an evolution operator for a map to its traces (7.7), that is, periodic orbits:

$$\begin{aligned}\det(1 - z\mathcal{L}) &= \exp\left(-\sum_n^{\infty} \frac{z^n}{n}\mathrm{tr}\,\mathcal{L}^n\right) \\ &= \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{z^{n_p r} e^{r\beta\cdot A_p}}{\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right|}\right)\,.\end{aligned} \tag{8.3}$$

Going the other way, the trace formula (7.9) can be recovered from the spectral determinant by taking a derivative

$$\mathrm{tr}\,\frac{z\mathcal{L}}{1 - z\mathcal{L}} = -z\frac{d}{dz}\ln\det(1 - z\mathcal{L})\,. \tag{8.4}$$

fast track: sect. 8.2, p. 149

8.1.1 Spectral determinants of transfer operators

For a piecewise-linear map (5.11) with a finite Markov partition, an explicit formula for the spectral determinant follows by substituting the trace

formula (7.12) into (8.3):

$$\det(1 - z\mathcal{L}) = \prod_{k=0}^{\infty}\left(1 - \frac{t_0}{\Lambda_0^k} - \frac{t_1}{\Lambda_1^k}\right) , \tag{8.5}$$

where $t_s = z/|\Lambda_s|$. The eigenvalues are - as they should be - (7.11), the ones that we already determined from the trace formula (7.9).

The exponential spacing of eigenvalues guarantees that the spectral determinant (8.5) is an entire function. It is this property that will generalize to piecewise smooth flows with finite Markov parititions, and single out spectral determinants rather than the trace formulas or dynamical zeta functions as the tool of choice for evaluation of spectra.

8.2 Spectral determinant for flows

> ... an analogue of the [Artin-Mazur] zeta function for diffeomorphisms seems quite remote for flows. However we will mention a wild idea in this direction. [···] define $l(\gamma)$ to be the minimal period of γ [···] then define formally (another zeta function!) $Z(s)$ to be the infinite product
>
> $$Z(s) = \prod_{\gamma\in\Gamma}\prod_{k=0}^{\infty}\left(1 - [\exp l(\gamma)]^{-s-k}\right) .$$
>
> Stephen Smale, *Differentiable Dynamical Systems*

We write the formula for the spectral determinant for flows by analogy to (8.3)

$$\det(s - \mathcal{A}) = \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{r(\beta\cdot A_p - sT_p)}}{\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right|}\right) , \tag{8.6}$$

and then check that the trace formula (7.19) is the logarithmic derivative of the spectral determinant so defined

$$\operatorname{tr}\frac{1}{s - \mathcal{A}} = \frac{d}{ds}\ln\det(s - \mathcal{A}) . \tag{8.7}$$

To recover $\det(s - \mathcal{A})$ integrate both sides $\int_{s_0}^{s} ds$. With z set to $z = e^{-s}$ as in (7.20), the spectral determinant (8.6) has the same form for both maps and flows.

We shall refer to (8.6) as *spectral determinant*, as the spectrum of the operator $\mathcal{A}$ is given by the zeros of

$$\det(s-\mathcal{A}) = 0\,. \tag{8.8}$$

We now note that the r sum in (8.6) 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 $\mathbf{J}_p$ be the p-cycle $[d\times d]$ transverse Jacobian matrix, with eigenvalues $\Lambda_{p,1}$, $\Lambda_{p,2}$, ..., $\Lambda_{p,d}$. Expanding $1/(1-1/\Lambda_{p,e})$, $1/(1-\Lambda_{p,c})$ in (7.4) as geometric series, substituting back into (8.6), and resumming the logarithms, we find that the spectral determinant is formally given by the infinite product

$$\begin{aligned}
\det(s-\mathcal{A}) &= \prod_{k_1=0}^{\infty}\cdots\prod_{l_c=0}^{\infty}\frac{1}{\zeta_{k_1\cdots l_c}} \\
1/\zeta_{k_1\cdots l_c} &= \prod_p\left(1-t_p\frac{\Lambda_{p,e+1}^{l_1}\Lambda_{p,e+2}^{l_2}\cdots\Lambda_{p,d}^{l_c}}{\Lambda_{p,1}^{k_1}\Lambda_{p,2}^{k_2}\cdots\Lambda_{p,e}^{k_e}}\right) \qquad (8.9)\\
t_p &= t_p(z,s,\beta) = \frac{1}{|\Lambda_p|}e^{\beta\cdot A_p - sT_p}z^{n_p}\,. \qquad (8.10)
\end{aligned}$$

Here we have inserted a topological cycle length weigth z^{n_p} for reasons which will become apparent in chapter 13; eventually we shall set $z=1$. The observable whose average we wish to compute contributes through the A_p term, which is the p cycle average of the multiplicative weight $e^{A^t(x)}$. By its definition (6.1), for maps the weight is a product along the cycle points

$$e^{A_p} = \prod_{j=0}^{n_p-1} e^{a(f^j(x_p))}\,,$$

and for the flows the weight is an exponential of the integral (6.5) 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 Jacobian matrix of sect. 4.1.

Now we are finally poised to deal with the problem posed at the beginning of chapter 7; how do we actually evaluate the averages introduced in sect. 6.1? The

eigenvalues of the dynamical averaging evolution operator are given by the values of s for which the spectral determinant (8.6) of the evolution operator (6.19) vanishes. If we can compute the leading eigenvalue $s_0(\beta)$ and its derivatives, we are done. Unfortunately, the infinite product formula (8.9) 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 cycle formulas into a tractable form. This we shall accomplish in chapter 13, but this point in the narrative is a natural point to introduce a still another variant of a determinant, the dynamical zeta function.

8.3 Dynamical zeta functions

It follows from sect. 7.1.1 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 Jacobian matrix $\mathbf{J}_p$. With this replacement the spectral determinant (8.6) is replaced by the *dynamical zeta function*

$$1/\zeta \;=\; \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{1}{r} t_p^r\right) \tag{8.11}$$

that we have already derived heuristically in sect. 1.4.2. Resumming the logarithms using $\sum_r t_p^r/r = -\ln(1-t_p)$ we obtain the *Euler product rep.* of the dynamical zeta function:

$$1/\zeta \;=\; \prod_p (1-t_p)\ . \tag{8.12}$$

For reasons of economy of the notation, we shall usually omit the explicit dependence of $1/\zeta$, t_p on z, s, β whenever the dependence is clear from the context.

The approximate trace formula (7.24) plays the same role vis-a-vis the dynamical zeta function

$$\Gamma(s) = \frac{d}{ds} \ln \zeta^{-1} = \sum_p \frac{T_p t_p}{1-t_p}\,, \tag{8.13}$$

as the exact trace formula (7.19) plays vis-a-vis the spectral determinant (8.6), see (8.7). The heuristically derived dynamical zeta function of sect. 1.4.2 now re-emerges as the $1/\zeta_{0\cdots0}(z)$ part of the *exact* spectral determinant; other factors in the infinite product (8.9) affect the non-leading eigenvalues of $\mathcal{L}$.

To summarize: the dynamical zeta function (8.12) associated with the flow $f^t(x)$ is defined as the product over all prime cycles p. T_p, n_p and Λ_p are the period, topological length and stability of prime cycle p, A_p is the integrated observable $a(x)$ evaluated on a single traversal of cycle p (see (6.5)), s is a variable dual to the time t, z is dual to the discrete "topological" time n, and $t_p(z, s, \beta)$ is 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

$$1/\zeta(s) = 0 \tag{8.14}$$

vanishes at s equal to s_0, the leading eigenvalue of $\mathcal{L}^t = e^{t\mathcal{A}}$, and often the leading eigenvalue is all that is needed in applications. The above 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 13, p. 293

8.3.1 A contour integral formulation

The following observation is sometimes useful, in particular when the zeta functions have richer analytic structure than just zeros and poles, as in the case of intermittency (chapter 16): Γ_n, the trace sum (7.22), can be expressed in terms of the dynamical zeta function (8.12)

$$1/\zeta(z) = \prod_p \left(1 - \frac{z^{n_p}}{|\Lambda_p|}\right) \quad . \tag{8.15}$$

as a contour integral

$$\Gamma_n = \frac{1}{2\pi i} \oint_{\gamma_r^-} z^{-n} \left(\frac{d}{dz} \log \zeta^{-1}(z)\right) dz \quad , \tag{8.16}$$

8.6
on p. 165

where a small contour γ_r^- encircles the origin in negative (clockwise) direction. If the contour is small enough, that is it lies inside the unit circle $|z| = 1$, we may

Figure 8.1: The survival probability Γ_n can be split into contributions from poles (x) and zeros (o) between the small and the large circle and a contribution from the large circle.

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 (7.22) is recovered. For hyperbolic maps, cycle expansion or other techniques provide an analytic extension of the dynamical zeta function beyond the leading zero; we may therefore deform the orignal contour into a larger circle with radius R which encircles both poles and zeros of $\zeta^{-1}(z)$, see fig. 16.5. Residue calculus turns this into a sum over the zeros z_α and poles z_β of the dynamical zeta function, that is

$$\Gamma_n = \sum_{|z_\alpha|<R}^{\text{zeros}} \frac{1}{z_\alpha^n} - \sum_{|z_\beta|<R}^{\text{poles}} \frac{1}{z_\beta^n} + \frac{1}{2\pi i}\oint_{\gamma_R^-} dz\, z^{-n} \frac{d}{dz}\log\zeta^{-1}, \tag{8.17}$$

where the last term gives a contribution from a large circle γ_R^-. We thus find exponential decay of Γ_n dominated by the leading zero or pole of $\zeta^{-1}(z)$.

8.3.2 Dynamical zeta functions for transfer operators

Ruelle's original dynamical zeta function was a generalization of the topological zeta function (11.20) that we shall discuss in chapter 11 to a function that assigns different weights to different cycles:

$$\zeta(z) = \exp\sum_{n=1}^{\infty}\frac{z^n}{n}\left(\sum_{x_i\in \mathrm{Fix} f^n} \mathrm{tr}\prod_{j=0}^{n-1} g(f^j(x_i))\right) .$$

7.2 on p. 146

Here the sum goes over all periodic points x_i of period n, and $g(x)$ is any (matrix valued) weighting function, with weight evaluated multiplicatively along the trajectory of x_i.

By the chain rule the stability of any n-cycle of a 1-d map factorizes as $\Lambda_p = \prod_{j=1}^{n} f'(x_i)$, so the 1-$d$ map cycle stability is the simplest example of a multiplicative cycle weight $g(x_i) = f'(x_i)$, and indeed - via the Perron-Frobenius evolution operator (5.9) - the historical motivation for Ruelle's more abstract construction.

In particular, for a piecewise-linear map with a finite Markov partition, the dynamical zeta function is given by a finite polynomials, a straightforward generalization of determinant of the topological transition matrix (10.2). As explained in sect. 11.3, 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. We illustrate this by the piecewise linear repeller (5.11). Due to the piecewise linearity, the stability of any n-cycle factorizes as $\Lambda_{s_1 s_2 \ldots s_n} = \Lambda_0^m \Lambda_1^{n-m}$, where m is total number of times letter $s_j = 0$ appears in the p symbol sequence, so the traces in the sum (7.24) are of a particularly simple form

$$\mathrm{tr}\, T^n = \Gamma_n = \left(\frac{1}{|\Lambda_0|} + \frac{1}{|\Lambda_1|} \right)^n .$$

8.2 The dynamical zeta function (8.11) evaluated by resumming the traces
on p. 164

$$1/\zeta(z) = 1 - z/|\Lambda_0| - z/|\Lambda_1| \tag{8.18}$$

is indeed the determinant $\det(1 - zT)$ of the transfer operator (5.13), almost as simple as the topological zeta function (11.24). 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 (see sect. 10.2).

We see that the exponential proliferation of cycles so dreaded by quantum chaoticists 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 13 how to exploit systematically this redundancy.

8.4 False zeros

Compare (8.18) with the Euler product (8.12). For simplicity take the two scales equal, $|\Lambda_0| = |\Lambda_1| = e^{\lambda}$. 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 (8.12) vanishes whenever one of its factors vanishes. If that were true, each factor $(1 - z^{n_p}/|\Lambda_p|)$ would yield

$$0 = 1 - e^{n_p(\gamma - \lambda_p)}, \tag{8.19}$$

that is the escape rate γ would equal the stability exponent of a repulsive fixed point. False! The exponentially growing number of cycles with growing period conspires to shift the zeros of the infinite product. The correct formula follows from (8.18)

$$0 = 1 - e^{\gamma - \lambda + h}, \qquad h = \ln 2. \tag{8.20}$$

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. The physical interpretation is that the escape induced by repulsion by each unstable fixed point is diminished by the rate of backscatter from other repelling segments, that is the entropy h; the positive entropy of orbits of the same stability shifts the "false zeros" $z = e^{\lambda_p}$ of the Euler product (8.12) to the true zero $z = e^{\lambda - h}$.

8.5 More examples of spectral determinants

For expanding 1-d mappings the spectral determinant (8.9) takes form

$$\det(s - \mathcal{A}) = \prod_p \prod_{k=0}^{\infty} \left(1 - t_p/\Lambda_p^k\right), \qquad t_p = \frac{e^{\beta A_p - sT_p}}{|\Lambda_p|} z^{n_p}. \tag{8.21}$$

For a periodic orbit of a 2-dimensional hyperbolic Hamiltonian flow with one expanding transverse eigenvalue Λ, $|\Lambda| > 1$, and one contracting transverse eigenvalue $1/\Lambda$, the weight in (7.4) is expanded as follows:

$$\frac{1}{\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right|} = \frac{1}{|\Lambda|^r (1 - 1/\Lambda_p^r)^2} = \frac{1}{|\Lambda|^r} \sum_{k=0}^{\infty} \frac{k+1}{\Lambda_p^{kr}}. \tag{8.22}$$

The spectral determinant exponent can be resummed,

$$-\sum_{r=1}^{\infty}\frac{1}{r}\frac{e^{(\beta A_p - sT_p)r}}{\left|\det\left(\mathbf{1}-\mathbf{J}_p^r\right)\right|} = \sum_{k=0}^{\infty}(k+1)\log\left(1-\frac{e^{\beta A_p - sT_p}}{|\Lambda_p|\Lambda_p^k}\right) ,$$

and the spectral determinant for a 2-dimensional hyperbolic Hamiltonian flow rewritten as an infinite product over prime cycles

$$\det(s-\mathcal{A}) = \prod_p \prod_{k=0}^{\infty}\left(1-t_p/\Lambda_p^k\right)^{k+1} . \tag{8.23}$$

9.4 on p. 194

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. We use z as a formal parameter which keeps track of the topological cycle lengths, to assist us in expanding zeta functions and determinants, then set it to $z = 1$ in calculations.

8.5.1 Spectral determinants *vs.* dynamical zeta functions

In sect. 7.2 we derived the dynamical zeta function as an approximation to the spectral determinant. Here we relate dynamical zeta functions to the 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

$$1 = \frac{1}{\det(1-\mathbf{J})}\sum_{k=0}^{d}(-1)^k \mathrm{tr}\left(\wedge^k \mathbf{J}\right) , \tag{8.24}$$

inserted into the exponential representation (8.11) of the dynamical zeta function, relates the dynamical zeta function to *weighted* spectral determinants. For 1-d maps the identity

$$1 = \frac{1}{(1-1/\Lambda)} - \frac{1}{\Lambda}\frac{1}{(1-1/\Lambda)}$$

substituted into (8.11) yields an expression for the dynamical zeta function for 1-d maps as a ratio of two spectral determinants

$$1/\zeta = \frac{\det(1-\mathcal{L})}{\det(1-\mathcal{L}_{(1)})} \tag{8.25}$$

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 9, this establishes that for nice hyperbolic flows $1/\zeta$ is meromorphic, with poles given by the zeros of $\det(1-\mathcal{L}_{(1)})$. The dynamical zeta function and the spectral determinant have the same zeros - only in exceptional circumstances some zeros of $\det(1-\mathcal{L}_{(1)})$ might be cancelled by coincident zeros of $\det(1-\mathcal{L}_{(1)})$. Hence even though we have derived the dynamical zeta function in sect. 8.3 as an "approximation" to the spectral determinant, the two contain the same spectral information.

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

$$1/\zeta = \frac{\det(1-\mathcal{L})\det(1-\mathcal{L}_{(2)})}{\det(1-\mathcal{L}_{(1)})}\,. \tag{8.26}$$

This establishes that for nice hyperbolic flows dynamical zeta function is meromorphic in 2-d.

8.5.2 Dynamical zeta functions for 2-d Hamiltonian flows

The relation (8.26) 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 (8.11) of $1/\zeta_k$, and obtain

$$1/\zeta_k = \frac{F_k F_{k+2}}{F_{k+1}^2}\,. \tag{8.27}$$

Even though we have no guarantee that F_k are entire, we do know (by arguments explained in sect. ?!) that the upper bound on the leading zeros of F_{k+1} lies strictly below the leading zeros of F_k, and therefore we expect that for 2-dimensional Hamiltonian flows the dynamical zeta function $1/\zeta_k$ has generically a *double* leading pole coinciding with the leading zero of the F_{k+1} spectral determinant. This might fail if the poles and leading eigenvalues come in wrong order, but we have not encountered such situation in our numerical investigations. This result can also be stated as follows: the theorem that establishes that the spectral determinant (8.23) is entire, implies that the poles in $1/\zeta_k$ must have right multiplicities in order that they be cancelled in the $F = \prod 1/\zeta_k^{k+1}$ product.

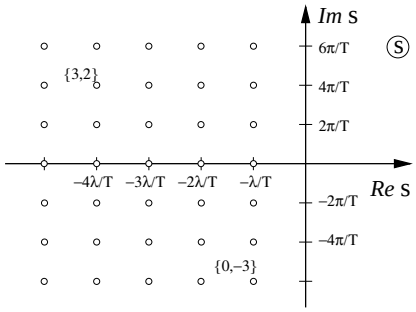


Figure 8.2: The classical resonances $\alpha = \{k, n\}$ for a 2-disk game of pinball, equation (8.28).

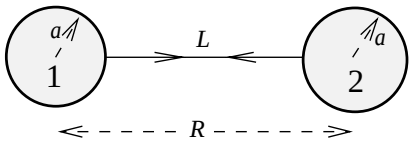


Figure 8.3: A game of pinball consisting of two disks of equal size in a plane, with its only periodic orbit. (A. Wirzba)

8.6 All too many eigenvalues?

What does the 2-dimensional hyperbolic Hamiltonian flow spectral determinant (8.23) tell us? Consider one of the simplest conceivable hyperbolic flows: the game of pinball of fig. 8.3 consisting of two disks of equal size in a plane. There is only one periodic orbit, with the period T and the expanding eigenvalue Λ is given by elementary considerations (see exercise 4.4), and the resonances $\det(s_\alpha - \mathcal{A}) = 0$, $\alpha = \{k, n\}$ plotted in fig. 8.2

$$s_\alpha = -(k+1)\lambda + n\frac{2\pi i}{T}\,, \quad n \in \mathbb{Z}\,,\ k \in \mathbb{Z}_+\,, \quad \text{multiplicity } k+1\ , \tag{8.28}$$

can be read off the spectral determinant (8.23) for a single unstable cycle:

$$\det(s - \mathcal{A}) = \prod_{k=0}^{\infty}\left(1 - e^{-sT}/|\Lambda|\Lambda^k\right)^{k+1}\ . \tag{8.29}$$

In the above $\lambda = \ln|\Lambda|/T$ is the cycle Lyapunov exponent. For an open system, the real part of the eigenvalue s_α gives the decay rate of αth eigenstate, and the imaginary part gives the "node number" of the eigenstate. The negative real part of s_α indicates that the resonance is unstable, and the decay rate in this simple case (zero entropy) equals to the cycle Lyapunov exponent.

Fast 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 (7.19), the trace formula for flows:

$$\begin{aligned}
\sum_{\alpha=0}^{\infty} e^{s_\alpha t} &= \sum_{k=0}^{\infty}\sum_{n=-\infty}^{\infty} (k+1)e^{(k+1)\lambda t + i2\pi n t/T} \\
&= \sum_{k=0}^{\infty}(k+1)\left(\frac{1}{|\Lambda|\Lambda^k}\right)^{t/T} \sum_{n=-\infty}^{\infty} e^{i2\pi n/T} \\
&= \sum_{k=0}^{\infty}\frac{k+1}{|\Lambda|^r\Lambda^{kr}} \sum_{r=-\infty}^{\infty} \delta(r-t/T) \\
&= T\sum_{r=-\infty}^{\infty} \frac{\delta(t-rT)}{|\Lambda|(1-1/\Lambda^r)^2}
\end{aligned} \tag{8.30}$$

So the two sides of the trace formula (7.19) check. The formula is fine for $t > 0$; for $t \to 0_+$ both sides are divergent and need regularization.

The reason why such sums do not occur for maps is that for discrete time we work in the variable $z = e^s$, an infinite strip along $\mathrm{Im}\, s$ maps into an anulus in the complex z plane, and the Dirac delta sum in the above is replaced by the Kronecker delta sum in (7.7). In case at hand there is only one time scale T, and we could as well replace s by variable $z = e^{-s/T}$. In general the flow has a continuum of cycle periods, and the resonance arrays are more irregular, *cf.* fig. 13.1.

Commentary

Remark **8.1** Piecewise monotone maps. A partial list of cases for which the transfer operator is well defined: expanding Hölder case, weighted subshifts of finite type, expanding differentiable case, see Bowen [13]: expanding holomorphic case, see Ruelle [9]; piecewise monotone maps of the interval, see Hofbauer and Keller [14] and Baladi and Keller [17].

Remark **8.2** Smale's wild idea. Smale's wild idea quoted on page 149 was technically wrong because 1) the Selberg zeta 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 is dimensionally 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 normalizing

the time so that $\lambda = 1$ the factors e^{-sT_p}/Λ_p^k in (8.9) simplify to $s^{-(s+k)T_p}$, as intuited in Smale's wild idea quoted on page 149 (where $l(\gamma)$ is the cycle period denoted here by T_p). Nevertheless, Smale's intuition was remarkably on the target.

Remark **8.3** Is this a generalization of the Fourier analysis? The Fourier analysis is a theory of the space $\leftrightarrow$ eignfunctions duality for dynamics on a circle. The sense in which the periodic orbit theory is the generalization of the Fourier analysis to nonlinear flows is discussed in ref. [4], a very readable introduction to the Selberg Zeta function.

Remark **8.4** Zeta functions, antecedents. For a function to be deserving of the appellation "zeta function", one expects it to have an Euler product (8.12) representation, and perhaps also satisfy a functional equation. Various kinds of zeta functions are reviewed in refs. [8, 9, 10]. Historical antecedents of the dynamical zeta function are the fixed-point counting functions introduced by Weil [11], Lefschetz [12] and Artin and Mazur [13], and the determinants of transfer operators of statistical mechanics [14].

In his review article Smale [12] 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 study of classical and quantal spectra are the spectral determinants. Ruelle had derived the relation (8.3) between spectral determinants and dynamical zeta functions, but as he was motivated by the Artin-Mazur zeta function (11.20) and the statistical mechanics analogy, he did not consider the spectral determinant a more natural object than the dynamical zeta function. This has been put right in papers on "flat traces" [22, 27].

The nomenclature has not settled down yet; what we call evolution operators here is called transfer operators [16], Perron-Frobenius operators [6] and/or Ruelle-Araki operators elsewhere. Here we refer to kernels such as (6.19) as evolution operators. We follow Ruelle in usage of the term "dynamical zeta function", but elsewhere in the literature function (8.12) is often called the Ruelle zeta function. Ruelle [18] points out 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 [4], functional determinants, Fredholm determinants, or even - to maximize confusion - dynamical zeta functions [**?**]. A Fredholm determinant is a notion that applies only to the trace class operators - as we consider here a somewhat wider class of operators, we prefer to refer to their determinants losely as "spectral determinants".

Résumé

The spectral problem is now recast into a problem of determining zeros of either the *spectral determinant*

$$\det(s-\mathcal{A}) = \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{(\beta\cdot A_p - sT_p)r}}{\left|\det\left(\mathbf{1}-\mathbf{J}_p^r\right)\right|}\right),$$

or the leading zeros of the *dynamical zeta function*

$$1/\zeta = \prod_p (1-t_p), \qquad t_p = \frac{1}{|\Lambda_p|} e^{\beta\cdot A_p - sT_p}.$$

The spectral determinant is the tool of choice in actual calculations, as it has superior convergence properties (this will be discussed in chapter 9 and is illustrated, for example, by table 13.2). In practice both spectral determinants and dynamical zeta functions are preferable to trace formulas because they yield the eigenvalues more readily; the main difference is that while a trace diverges at an eigenvalue and requires extrapolation methods, determinants vanish at s corresponding to an eigenvalue s_α, and are analytic in s in an open neighborhood of s_α.

The critical step in the derivation of the periodic orbit formulas for spectral determinants and dynamical zeta functions is the hyperbolicity assumption, that is the assumption that all cycle stability eigenvalues are bounded away from unity, $|\Lambda_{p,i}| \neq 1$. By dropping the prefactors in (1.4), we have given up on any possibility of recovering the precise distribution of starting 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 13) is that the infinite time behavior of an unstable system turns out to be as easy to determine as its short time behavior.

References

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

[8.2] D. Ruelle, *Bull. Amer. Math. Soc.* **78**, 988 (1972)

[8.3] M. Pollicott, *Invent. Math.* **85**, 147 (1986).

[8.4] H.P. McKean, *Comm. Pure and Appl. Math.* **25** , 225 (1972); **27**, 134 (1974).

[8.5] W. Parry and M. Pollicott, *Ann. Math.* **118**, 573 (1983).

[8.6] 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).

[8.7] P. Cvitanović, P.E. Rosenqvist, H.H. Rugh, and G. Vattay, *CHAOS* **3**, 619 (1993).

[8.8] 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.

[8.9] Kiyosi Itô, ed., *Encyclopedic Dictionary of Mathematics*, (MIT Press, Cambridge, 1987).

[8.10] N.E. Hurt, "Zeta functions and periodic orbit theory: A review", *Results in Mathematics* **23**, 55 (Birkhäuser, Basel 1993).

[8.11] A. Weil, *"Numbers of solutions of equations in finite fields"*, *Bull. Am. Math. Soc.* **55**, 497 (1949).

[8.12] D. Fried, "Lefschetz formula for flows", *The Lefschetz centennial conference*, *Contemp. Math.* **58**, 19 (1987).

[8.13] E. Artin and B. Mazur, *Annals. Math.* **81**, 82 (1965)

[8.14] F. Hofbauer and G. Keller, *"Ergodic properties of invariant measures for piecewise monotonic transformations"*, *Math. Z.* **180**, 119 (1982).

[8.15] G. Keller, *"On the rate of convergence to equilibrium in one-dimensional systems"*, *Comm. Math. Phys.* **96**, 181 (1984).

[8.16] F. Hofbauer and G. Keller, *"Zeta-functions and transfer-operators for piecewise linear transformations"*, *J. reine angew. Math.* **352**, 100 (1984).

[8.17] V. Baladi and G. Keller, *"Zeta functions and transfer operators for piecewise monotone transformations"*, *Comm. Math. Phys.* **127**, 459 (1990).

Exercises

8.1 Escape rate for a 1-d repeller, numerically. Consider the quadratic map

$$f(x) = Ax(1-x) \tag{8.31}$$

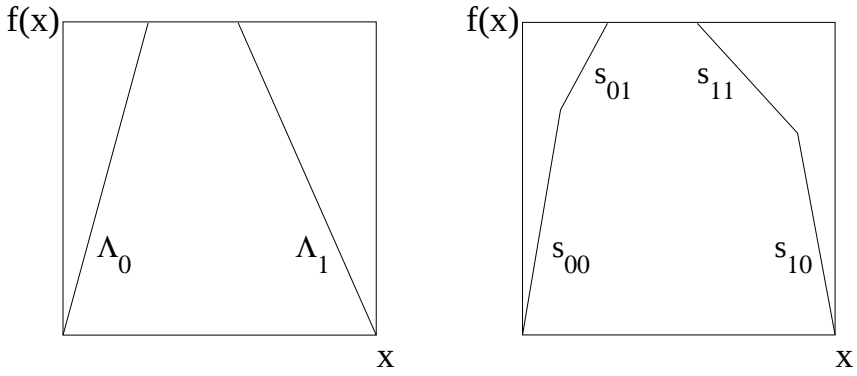
on the unit interval. The trajectory of a point starting in the unit interval either stays in the interval forever or after some iterate leaves the interval and diverges to minus infinity. Estimate numerically the escape rate (14.8), the rate of exponential decay of the measure of points remaining in 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 13.2.

8.2 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 (5.11) with the left branch slope Λ_0, the right branch slope Λ_1.



(b) What if there are four different slopes s_{00}, s_{01}, s_{10}, and s_{11} instead of just two, with the preimages of the gap adjusted so that junctions of branches s_{00}, s_{01} and s_{11}, s_{10} map in the gap in one iteration? What would the dynamical zeta function be?

8.3 Zeros of infinite products. Determination of the quantities of interest by periodic orbits involves working with infinite product formulas.

(a) Consider the infinite product

$$F(z) = \prod_{k=0}^{\infty}(1 + f_k(z))$$

where the functions f_k are "sufficiently nice." This infinite product can be converted into an infinite sum by the use of a logarithm. Use the properties of infinite sums to develop a sensible definition of infinite products.

(b) If $z_{\rm root}$ is a root of the function F, show that the infinite product diverges when evaluated at $z_{\rm root}$.

(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)

8.4 Dynamical zeta functions as ratios of spectral determinants. (medium) Show that the zeta function

$$1/\zeta(z) = \exp\left(-\sum_{p}\sum_{r>0}\frac{1}{r}\frac{z^{n_p}}{|\Lambda_p|^r}\right)$$

can be written as the ratio $1/\zeta(z) = \frac{\det(1-z\mathcal{L}_{(0)})}{\det(1-z\mathcal{L}_{(1)})}$,
where $\det(1 - z\mathcal{L}_{(s)}) = \prod_{p,k}(1 - z^{n_p}/|\Lambda_p|\Lambda_p^{k+s})$.

8.5 Escape rate for the Ulam map. (medium) We will try and compute the escape rate for the Ulam map (12.28)

$$f(x) = 4x(1 - x),$$

using 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.

8.6 Contour integral for survival probability. Perform explicitly the contour integral appearing in (8.16).

8.7 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 (5.41)

$$1/\zeta(z) = \prod_{p \in P} \left(1 - \frac{z^{n_p}}{|\Lambda_p|}\right) .$$

What are its roots? Do they agree with those computed in exercise 5.7?

8.8 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 two-dimensional Hamiltonian map and using the equality

$$1 = \frac{1}{(1-1/\Lambda)^2}(1 - 2/\Lambda + 1/\Lambda^2) ,$$

show that $1/\zeta = \frac{\det(1-\mathcal{L})\det(1-\mathcal{L}_{(2)})}{\det(1-\mathcal{L}_{(1)})^2}$. 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.

8.9 Riemann ζ function. The Riemann ζ 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 8.12 is called a "zeta" function because it shares the form of the Euler product representation with the Riemann zeta function.

(b) (Not trivial:) For which complex values of s is the Riemann zeta sum convergent?

(c) Are the zeros of the terms in the product, $s = -\ln p$, also the zeros of the Riemann ζ function? If not, why not?

8.10 <u>Finite truncations.</u> (easy) Suppose we have a one-dimensional system with complete binary dynamics, where the stability of each orbit is given by a simple multiplicative rule:

$$\Lambda_p = \Lambda_0^{n_{p,0}} \Lambda_1^{n_{p,1}}\,, \qquad n_{p,0} = \#0 \ in\ p\,,\ n_{p,1} = \#1 \ in\ p\,,$$

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 (**??**), with the right weights.

(b) Compute the finite p truncations of the cycle expansion, that is take the product only over the p up to given length with $n_p \leq N$, and expand as a series in z

$$\prod_p \left(1 - \frac{z^{n_p}}{|\Lambda_p|}\right)\ .$$

Do they agree? If not, how does the disagreement depend on the truncation length N?

8.11 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. [8] yields $\gamma = .410\ldots$.

Chapter 9

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)

The trace formulas and spectral determinants work well, sometimes very well indeed. The question is: why? The heuristic manipulations of chapter 7 were naive and reckless, as we are facing infinite-dimensional vector spaces and singular integral kernels.

In this chapter we outline some of the ingredients in the proofs that put the above trace and determinant formulas on solid mathematical footing. This requires taking a closer look at the Perron-Frobenius operator from a mathematical point of view, since up to now we have talked about eigenvalues without any reference to an underlying function space. In sect. 9.1 we show, by a simple example, that the spectrum is quite sensitive to the regularity properties of the functions considered, so what we referred to as *the set of eigenvalues* acquires a meaning only if the functional setting is properly tuned: this sets the stage for a discussion of analyticity properties mentioned in chapter 8. The program is enunciated in sect. 9.2, with the focus on expanding maps. In sect. 9.3 we concentrate on piecewise real-analytic maps acting on appropriate densities. For expanding and hyperbolic flows 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 being entire functions in the complex s plane.

This chapter is not meant to provide an exhaustive review of rigorous results about properties of the Perron-Frobenius operator or analyticity results of spec-

tral determinants or dynamical zeta functions (see remark 9.5), but rather to point out that heuristic considerations about traces and determinant can be put on firmer bases, under suitable hypotheses, and the mathematics behind this construction is both hard and profound.

If you are primarily interested in physical applications of periodic orbit theory, you should probably skip this chapter on the first reading.

fast track:
chapter 14, p. 319

9.1 The simplest of spectral determinants: A single fixed point

In order to get some feeling for the determinants defined so formally in sect. 8.2, 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 (5.10) is

$$\mathcal{L}\phi(y) = \int dx\, \delta(y - \Lambda x)\, \phi(x) = \frac{1}{|\Lambda|}\phi(y/\Lambda)\,. \tag{9.1}$$

From this one immediately gets that the monomials y^n are eigenfunctions:

$$\mathcal{L}y^n = \frac{1}{|\Lambda|\Lambda^n} y^n\,, \qquad n = 0, 1, 2, \ldots \tag{9.2}$$

We note that the eigenvalues Λ^{-n-1} fall off exponentially with n, and that the trace of $\mathcal{L}$ is

$$\operatorname{tr}\mathcal{L} = \frac{1}{|\Lambda|}\sum_{n=0}^{\infty}\Lambda^{-n} = \frac{1}{|\Lambda|(1-\Lambda^{-1})} = \frac{1}{|f(0)' - 1|}\,,$$

in agreement with (7.6). A similar result is easily obtained for powers of $\mathcal{L}$, and for the spectral determinant (8.3) one obtains:

$$\det(1 - z\mathcal{L}) = \prod_{k=0}^{\infty}\left(1 - \frac{z}{|\Lambda|\Lambda^k}\right) = \sum_{k=0}^{\infty} Q_k t^k\,, \qquad t = -z/|\Lambda|\,, \tag{9.3}$$

where the coefficients Q_k are given explicitly by the *Euler formula* 9.3 on p. 194

$$Q_k = \frac{1}{1-\Lambda^{-1}} \frac{\Lambda^{-1}}{1-\Lambda^{-2}} \cdots \frac{\Lambda^{-k+1}}{1-\Lambda^{-k}} \quad . \tag{9.4}$$

(if you cannot figure out exercise 9.3 check the solutions on 702 for proofs of this formula).

Note that the coefficients Q_k decay asymptotically *faster* than exponentially, as $\Lambda^{-k(k-1)/2}$. As we shall see in sect. 9.3.1, these results carry over to any single-branch repeller. This super-exponential decay of Q_k ensures that for a repeller consisting of a single repelling point the spectral determinant (9.3) is *entire* in the complex z plane.

What is the meaning of (9.3)? It gives us an interpretation of the index k in the Selberg product representation of the spectral determinant (8.9): k labels the kth local fixed-point eigenvalue $1/|\Lambda|\Lambda^k$.

Now if the spectral determinant is entire, on the basis of (8.25) we get that the dynamical zeta function is a meromorphic function. These mathematical properties are of direct physical import: they guarantee that finite order estimates of zeroes of dynamical zeta functions and spectral determinants converge exponentially or super-exponentially to the exact values, and so the cycle expansions of chapter 13 represent a *true perturbative* approach to chaotic dynamics. To see how exponential convergence comes out of analytic properties we take the simplest possible model of a meromorphic function. Consider the function

$$h(z) = \frac{z-a}{z-b}$$

with a, b real and positive and $a < b$. Within the cycle $|z| < b$ we may represent h as a power series

$$h(z) = \sum_{k=0}^{\infty} \sigma_k z^k$$

where $\sigma_0 = a/b$ and higher order coefficients are given by $\sigma_j = (a-b)/b^{j+1}$ Now we take the truncation of order N of the power series

$$h_N(z) = \sum_{k=0}^{N} \sigma_k z^k = \frac{a}{b} + \frac{z(a-b)(1-z^N/b^N)}{b^2(1-z/b)} .$$

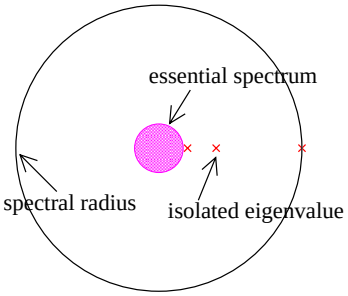


Figure 9.1: Spectrum for Perron-Frobenius operator in an extended function space: only a few isolated eigenvalues remain between the spectral radius and the essential spectral radius, bounding continuous spectrum

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 $h_N(a)$, which is of order $(a/b)^{N+1}$, and so finite order estimates indeed converge exponentially to the asymptotic value.

The discussion of our simple example confirms that our formal manipulations with traces and determinants are justified, namely the Perron-Frobenius operator has isolated eigenvalues: trace formulas are then explicitly verified, the spectral determinant is an analytic function whose zeroes yield the eigenvalues. Life is actually harder, as we may appreciate through the following considerations

- Our discussion tacitly assumed something that is physically entirely reasonable: our evolution operator is acting on the space of analytic functions, that is, 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: we might choose the function space $C^{k+\alpha}$, that is the space of k times differentiable functions whose k'th derivatives are Hölder continuous with an exponent $0 < \alpha \leq 1$: then every y^η with $\mathrm{Re}\,\eta > k$ is an eigenfunction of Perron-Frobenius operator and we have

9.1 on p. 194

$$\mathcal{L}y^\eta = \frac{1}{|\Lambda|\Lambda^\eta} y^\eta$$

  This spectrum is quite different from the analytic case: only a small number of isolated eigenvalues remain, enclosed between the unit disk and a smaller disk of radius $1/|\Lambda|^{k+1}$, (the so-called essential spectral radius) see fig. 9.1.

  In sect. 9.2 we will discuss this point further, with the aid of a less trivial one-dimensional example. We remark that our 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, which is necessary in order to prove asymptotic decay of correlations in $L^2(d\mu)$ setting.

9.2 on p. 194

- A deceptively innocent assumption hides behind many features discussed so far: that (9.1) maps a given function space into itself. This is strictly related to the *expanding* property of the map: if $f(x)$ is smooth in a domain D then $f(x/\Lambda)$ is smooth on a *larger* domain, provided $|\Lambda| > 1$. This is not obviously the case for hyperbolic systems in higher dimensions, and, as we shall see in sect. 9.3, extensions of the results obtained for expanding maps will be highly nontrivial,
- It is not a priori clear that the above analysis of a simple one-branch, one fixed point repeller can be extended to dynamical systems with a Cantor set infinity of periodic points: we show that next.

9.2 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

We now choose another paradigmatic example (the Bernoulli shift) and sketch the steps that lead to the proof that the corresponding spectral determinant is an entire function. Before doing that it is convenient to summarize a few facts about classical theory of integral equations.

9.2.1 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

$$\mathcal{K}\varphi(x) = \int_{\mathcal{Q}} dy\, \mathcal{K}(x,y)\varphi(y)\,, \tag{9.5}$$

and one is tempted to resort to the classical Fredholm theory in order to establish analyticity properties of spectral determinants. This path to enlightment is blocked by the singular nature of the kernel, which is a distribution, wheras the standard theory of integral equations usually concerns itself with regular kernels $\mathcal{K}(x,y) \in L^2(\mathcal{Q}^2)$. Here we briefly recall some steps of the Fredholm theory, before going to our major example in sect. 9.2.2.

The general form of Fredholm integral equations of the second kind is

$$\varphi(x) \;=\; \int_{\mathcal{Q}} dy\, \mathcal{K}(x,y)\varphi(y) \;+\; \xi(x) \tag{9.6}$$

where $\xi(x)$ is a given function in $L^2(\mathcal{Q})$ and the kernel $\mathcal{K}(x,y) \in L^2(\mathcal{Q}^2)$ (Hilbert-Schmidt condition). The natural object to study is then the linear integral operator (9.5), acting on the Hilbert space $L^2(\mathcal{Q})$: and 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 (see appendix J). A compact operator has the property that for every $\delta > 0$ only a *finite* number of linearly independent eigenvectors exist corresponding to eigenvalues whose absolute value exceeds δ, so we immediately realize (fig. 9.1) that much work is needed to bring Perron-Frobenius operators into this picture.

We rewrite (9.6) in the form

$$\mathcal{T}\varphi \;=\; \xi\,, \mathcal{T} \;=\; \mathbb{1} - \mathcal{K}\,. \tag{9.7}$$

The Fredholm alternative is now stated as follows: the equation $\mathcal{T}\varphi = \xi$ as a unique solution for every $\xi \in L^2(\mathcal{Q})$ **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 will be at most a denumerable set of λ for which the second part of Fredholm alternative holds: so apart from this set the inverse operator $(\mathbb{1} - \lambda\mathcal{T})^{-1}$ exists and is a bounded operator. When λ is sufficiently small we may look for a perturbative expression for such an inverse, as a geometric series

$$(\mathbb{1} - \lambda\mathcal{K})^{-1} \;=\; \mathbb{1} + \lambda\mathcal{K} + \lambda^2\mathcal{K}^2 + \cdots \;=\; \mathbb{1} + \lambda\mathcal{W}\,, \tag{9.8}$$

where each $\mathcal{K}^n$ is still a compact integral operator with kernel

$$\mathcal{K}^n(x,y) \;=\; \int_{\mathcal{Q}^{n-1}} dz_1 \,\ldots dz_{n-1}\, \mathcal{K}(x,z_1)\cdots\mathcal{K}(z_{n-1},y)\,,$$

and $\mathcal{W}$ is also compact, as it is given by the convergent sum of compact operators. The problem with (9.8) is that the series has a finite radius of convergence, while apart from a denumerable set of λ's the inverse operator is well defined. A fundamental result in the theory of integral equations consists in rewriting the resolving kernel $\mathcal{W}$ as a ratio of two *analytic* functions of λ

$$\mathcal{W}(x,y) \;=\; \frac{\mathcal{D}(x,y;\lambda)}{D(\lambda)}.$$

If we introduce the notation

$$\mathcal{K}\begin{pmatrix} x_1 \dots x_n \\ y_1 \dots y_n \end{pmatrix} = \begin{vmatrix} \mathcal{K}(x_1,y_1) & \dots & \mathcal{K}(x_1,y_n) \\ \dots & \dots & \dots \\ \mathcal{K}(x_n,y_1) & \dots & \mathcal{K}(x_n,y_n) \end{vmatrix}$$

we may write the explicit expressions

$$D(\lambda) \;=\; 1+\sum_{n=1}^{\infty}(-1)^n\,\frac{\lambda^n}{n!}\int_{\mathcal{Q}^n} dz_1\,\dots dz_n\,\mathcal{K}\begin{pmatrix} z_1 \dots z_n \\ z_1 \dots z_n \end{pmatrix} \;=\; \exp -\sum_{m=1}^{\infty}\frac{\lambda^m}{m}\,\mathrm{tr}\,\mathcal{K}^m \tag{9.9}$$

and

$$\mathcal{D}(x,y;\lambda) \;=\; \mathcal{K}\begin{pmatrix} x \\ y \end{pmatrix} + \sum_{n=1}^{\infty}(-1)^n\,\frac{\lambda^n}{n!}\int_{\mathcal{Q}^n} dz_1\dots dz_n\,\mathcal{K}\begin{pmatrix} x & z_1 & \dots & z_n \\ y & z_1 & \dots & z_n \end{pmatrix}$$

$D(\lambda)$ is known as the Fredholm determinant (see (8.24) and appendix J): it is an entire analytic function of λ, and $D(\lambda)=0$ only if $1/\lambda$ is an eigenvalue of $\mathcal{K}$.

We remark again that the whole theory is based on the compactness of the integral operator, that is on the functional properties (summability) of its kernel.

9.2.2 Bernoulli shift

Consider now the Bernoulli shift

$$x \mapsto 2x \bmod 1 \qquad\qquad x\in[0,1] \tag{9.10}$$

and look at spectral properties in appropriate function spaces. The Perron-Frobenius operator associated with this map is given by

$$\mathcal{L}h(y) = \frac{1}{2}h\left(\frac{y}{2}\right) + \frac{1}{2}h\left(\frac{y+1}{2}\right). \tag{9.11}$$

Spaces of summable functions as $L^1([0,1])$ or $L^2([0,1])$ are mapped into themselves by the Perron-Frobenius operator, and in both spaces the constant function $h \equiv 1$ is an eigenfunction with eigenvalue 1. This obviously does not exhaust the spectrum: if we focus our attention on $L^1([0,1])$ we also have a whole family of eigenfunctions, parametrized by complex θ with $\mathrm{Re}\,\theta > 0$. One verifies that

$$h_\theta(y) = \sum_{k\neq 0} \exp(2\pi i k y) \frac{1}{|k|^\theta} \tag{9.12}$$

is indeed an L^1-eigenfunction with (complex) eigenvalue $2^{-\theta}$, by varying θ one realizes that such eigenvalues fill out the entire unit disk. This casts out a 'spectral rug', also known as an essential spectrum, which hides all the finer details of the spectrum.

For a bounded linear operator $\mathcal{A}$ on a Banach space $\mathbf{\Omega}$, the spectral radius is the smallest positive number ρ_{spec} such the spectrum is inside the disk of radius ρ_{spec}, while the essential spectral radius is the smallest positive number ρ_{ess} such that outside the disk of radius ρ_{ess} the spectrum consists only of isolated eigenvalues of finite multiplicity (see fig. 9.1).

9.5 on p. 195

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. 9.1. We thus consider a smaller space, $C^{k+\alpha}$, 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 h_θ will cease to be eigenfunctions in the space $C^{k+\alpha}$. Only for integer valued $\theta = n$ the function h_n survives. In this way we arrive at a finite set of *isolated* eigenvalues $1, 2^{-1}, \cdots, 2^{-k}$, and an essential spectral radius $\rho_{ess} = 2^{-(k+\alpha)}$.

For this simple example, we may actually exactly write down the eigenfunctions: they coincide, up to a constant, with the Bernoulli polynomials $B_n(x)$. These polynomials are defined as successive derivatives of $te^{xt}/(e^t-1)$ evaluated at $t = 0$:

$$\mathcal{G}_t(x) = \frac{te^{xt}}{e^t-1} = \sum_{n=0}^{\infty} B_n(x)\frac{t^n}{n!}$$

so $B_0(x) = 1$, $B_1(x) = x - 1/2$, etc. .

If we let the Perron-Frobenius operator (9.11) act on the generating function $\mathcal{G}$, we get

$$\mathcal{L}\mathcal{G}_t(x) = \frac{1}{2}\left(\frac{te^{xt/2}}{e^t-1} + \frac{te^{t/2}e^{xt/2}}{e^t-1}\right) = \frac{t/2e^{xt/2}}{e^{t/2}-1} = \sum_{n=1}^{\infty} B_n(x)\frac{(t/2)^n}{n!}$$

it follows that each $B_n(x)$ is an eigenfunction of the Perron-Frobenius operator $\mathcal{L}$ with eigenvalue $1/2^n$. The persistence of a finite essential spectral radius would suggest that traces and determinants do not exist in this case either. The pleasant surprise is that they do, see remark 9.3.

We follow a simpler path and restrict the function space even further, namely to a space of analytic functions, i.e. for which the is convergent at each point of the interval $[0,1]$. With this choice things turn out easy and elegant. To be more specific let h 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 $R > 1$. In this the expansion property of the Bernoulli shift enter). If F denotes one of the inverse branches of the Bernoulli shift (**??**) the corresponding part of the Perron-Frobenius operator is given by $\mathcal{L}_F h(y) = s\ F'(y)\ h \circ F(y)$, using the Cauchy integral formula:

$$\mathcal{L}_F h(y) = s \oint_{\partial D} \frac{h(w)F'(y)}{w - F(y)} dw.$$

For reasons that will be made clear later we have introduced a sign $s = \pm 1$ of the given real branch $|F'(y)| = sF(y)$. For both branches of the Bernoulli shift s^2+1, 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 stays non-singular under this condition, i.e. $w - F(y) \neq 0$ whenever $w \in \partial D$ and $y \in \text{Cl } D$.

The problem is now reduced to the standard theory for Fredholm determinants. The integral kernel is no longer singular, traces and determinants are well-defined and we may even calculate the trace of $\mathcal{L}_F$ as a contour integral:

$$\text{tr } \mathcal{L}_F = \oint \frac{sF'(w)}{w - F(w)} dw.$$

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

9.6 on p. 195

$$\text{tr } \mathcal{L}_F = \frac{sF'(x^*)}{1 - F'(x^*)} = \frac{1}{|f'(x^*) - 1|},$$

justifies our previous ad hoc calculations of traces by means of Dirac delta functions. The full operator has two components corresponding to the two branches

og the . 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 in each , yielding the trace $(2^n-1)^{-1}$. Without further ado we substitute everything back and obtain the determinant,

$$\det(1-z\mathcal{L}) = \exp\left(-\sum_{n=1}\frac{z^n}{n}\,\frac{2^n}{2^n-1}\right) = \prod_{k=0}\left(1-\frac{z}{2^k}\right),$$

verifying the fact that the Bernoulli polynomials are eigenfunctions with eigenvalues $1/2^n$, $n = 0, 1, 2, \ldots$.

We worked out a very specific example, yet our conclusions can be generalized, provided a number of restrictive requirements are met by our dynamical systems:

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 away from 1),
4) the map (or the flow) is *real analytic*, that is it has a piecewise analytic continuation to a complex extension of the phase space.

These assumptions are romantic projections not lived up to 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 of changes do satisfy the above requirements.

Properties 1 and 2 enable us to represent the evolution operator as a matrix in an appropriate basis space; 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

$$\mathcal{L}^t(y,x) = |\Lambda^t(x)|^\beta \delta\big(y-f^t(x)\big)\ ,$$

where $\Lambda^t(x)$ is an eigenvalue of the Jacobian matrix transverse to the flow. Semiclassical quantum mechanics suggest operators of this form with $\beta = 1/2$, (see chapter 22). The problem with such operators is due to the fact that when considering the Jacobian matrices $\mathbf{J}_{ab} = \mathbf{J}_a\mathbf{J}_b$ 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 repeats of the same prime cycle p, so $\mathbf{J}_a\mathbf{J}_b = \mathbf{J}_p^{r_a+r_b}$). Consequently, this evolution operator is not multiplicative along the trajectory. The

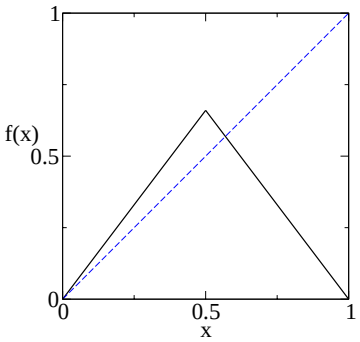


Figure 9.2: A (hyperbolic) tent map without a finite Markov partition.

theorems require that the evolution be represented as a matrix in an appropriate polynomial basis, and thus cannot be applied to non-multiplicative kernels, *that is*}. *kernels that do not satisfy the semi-group property* $\mathcal{L}^{t'} \circ \mathcal{L}^{t} = \mathcal{L}^{t'+t}$. *Cure for this problem in this particular case will be given in sect. G.1.*

Property 2 is violated by the 1-d tent map (see fig. 9.2)

$$f(x) = \alpha(1 \ - \ |1 - 2x|) \ , \quad 1/2 < \alpha < 1 \ .$$

All cycle eigenvalues are hyperbolic, but in general the critical point $x_c = 1/2$ *is not a pre-periodic point, there is no finite Markov partition and the symbolic dynamics does not have a finite grammar (see sect. 10.7 for definitions). In practice this means that while the leading eigenvalue of* $\mathcal{L}$ *might be computable, the rest of the spectrum is very hard to control; as the parameter* α *is varied, non-leading zeros of the spectral determinant move wildly about.*

Property 3 is violated by the map (see fig. 9.3)

$$f(x) = \begin{cases} x + 2x^2 & , \quad x \in I_0 = [0, \frac{1}{2}] \\ 2 - 2x & , \quad x \in I_1 = [\frac{1}{2}, 1] \end{cases} \ .$$

Here the interval $[0, 1]$ *has a Markov partition into the two subintervals* I_0 *and* I_1*;* f *is monotone on each. However, the fixed point at* $x = 0$ *has marginal stability* $\Lambda_0 = 1$*, and violates the 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 16.*

The property 4 is required as the heuristic approach of chapter 7 faces two major hurdles:

1. *The trace (7.7) is not well defined since the integral kernel is singular.*

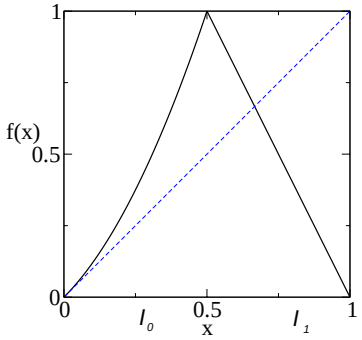


Figure 9.3: A Markov map with a marginal fixed point.

2. The existence and properties of eigenvalues are by no means clear.

Actually this property is quite restrictive, but we need it in the present approach, in order 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, $x \mapsto 2x \ mod\ 1$, $x \in R/Z$. Compared to the shift on the interval $[0,1]$ the only difference is that the endpoints 0 and 1 are now glued together. But since these endpoints are fixed points of the map the number of cycles of length n decreases by 1. The determinant becomes:

$$det(1 - z\mathcal{L}) = \exp\left(-\sum_{n=1} \frac{z^n}{n} \frac{2^n - 1}{2^n - 1}\right) = 1 - z. \tag{9.13}$$

The value $z = 1$ still comes from the constant eigenfunction but the Bernoulli polynomials no longer contribute to the spectrum (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.*

9.3 Hyperbolic maps

(H.H. Rugh)

Moving on to hyperbolic systems, one faces the following paradox: If f is an area-preserving hyperbolic and real-analytic map of e.g. a two dimensional torus then the Perron-Frobenius operator is clearly unitary on the space of L^2 functions. The spectrum is then confined to the unit-circle. On the other hand when we compute determinants we find eigenvalues scattered around inside the unit disk. Thinking back on our Bernoulli shift example 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 resort to analytic functions as before but in the stable direction one considers a 'dual space' of distributions on analytic functions. Such a space is neither included in nor does it include the L^2-space and we have thus resolved the paradox. But it still remains to be seen how traces and determinants are calculated.

First, let us consider the apparently trivial linear example ($0 < \lambda_s < 1$, $\Lambda_u > 1$):

$$f(z) = (f_1(z_1, z_2), f_2(z_1, z_2)) = (\lambda_s z_1, \Lambda_u z_2) \tag{9.14}$$

The function space, alluded to above, is then a mixture of Laurent series in the z_1 variable and analytic functions in the z_2 variable. Thus, one considers expansions in terms of $\varphi_{n_1,n_2}(z_1, z_2) = z_1^{-n_1-1} z_2^{n_2}$ with $n_1, n_2 = 0, 1, 2, \ldots$ If one looks at the corresponding Perron-Frobenius operator, one gets a simple generalization of the 1-d repeller:

$$\mathcal{L}h(z_1, z_2) = \frac{1}{\lambda_s \cdot \Lambda_u} h(z_1/\lambda_s, z_2/\Lambda_u) \tag{9.15}$$

The action of Perron-Frobenius operator on the basis functions yields

$$\mathcal{L}\varphi_{n_1,n_2}(z_1, z_2) \;=\; \frac{\lambda_s^{n_1}}{\Lambda_u^{1+n_2}} \varphi_{n_1,n_2}(z_1, z_2)$$

so that the above basis elements are eigenvectors with eigenvalues $\lambda_s^{n_1} \Lambda_u^{-n_2-1}$ and one verifies by an explicit calculation that the trace indeed equals $det(f' - \mathbf{1})^{-1} = (\Lambda_u - 1)^{-1}(1 - \lambda_s)^{-1}$.

This example is somewhat misleading, however, as we have made explicit use of an analytic 'splitting' into stable/unstable directions. For a more general hyperbolic map, if one attempts to make such a splitting it will not be analytic and the whole argument falls apart. Nevertheless, one may introduce 'almost' analytic splittings and write down a generalization of the above operator as follows (s is the signature of the derivative in the unstable direction):

$$\mathcal{L}h(z_1,z_2) = \oint\oint \frac{s\ h(w_1,w_2)}{(z_1 - f_1(w_1,w_2)(f_2(w_1,w_2) - z_2)} \frac{dw_1}{2\pi i}\frac{dw_2}{2\pi i}. \tag{9.16}$$

Here the 'function' h should belong to a space of functions analytic respectively outside *a disk and* inside *a disk in the first and the second coordinate and 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 but an exercise in multi-dimensional residue calculus to verify that for the above linear example this expression reduces to (9.15). Such operators form the building bricks in the calculation of traces and determinants and one is able to prove the following:*

Theorem: The spectral determinant for hyperbolic analytic maps is entire.

The proof, apart from the Markov property which is the same as for the purely expanding case, relies heavily on analyticity of the map in the explicit construction of the function space. As we have also seen in the previous example the basic idea is to view the hyperbolicity as a cross product of a contracting map in the forward time and another contracting map in the backward time. In this case the Markov property introduced above has to be elaborated a bit. Instead of dividing the phase 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 9.11) one may choose coordinate axes close to the stable/unstable manifolds of the map. With the phase 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 complex extension $D_i^h \times D_i^v$, with which the hyperbolicity condition (which at the same time guarantees the Markov property) can be formulated as follows:

Analytic hyperbolic property: Either $f(\mathcal{M}_i) \cap Int(\mathcal{M}_j) = \emptyset$, or for each pair $w_h \in Cl(D_i^h)$, $z_v \in Cl(D_j^v)$ there exist unique analytic functions of w_h, z_v: $w_v = w_v(w_h, z_v) \in Int(D_i^v)$, $z_h = z_h(w_h, z_v) \in Int(D_j^h)$, 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 fig. 9.4).

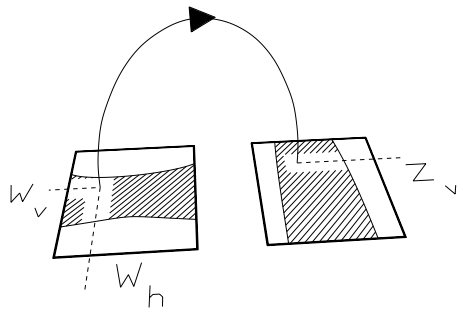


Figure 9.4: For an analytic hyperbolic map, specifying the contracting coordinate w_h at the initial rectangle and the expanding coordinate z_v at the image rectangle defines a unique trajectory between the two rectangles. In particular, w_v and z_h (not shown) are uniquely specified.

What this means for the iterated map is that one replaces coordinates z_h, z_v at time n by the contracting pair z_h, w_v, where w_v is the contracting coordinate at time $n+1$ for the 'partial' inverse map.

In two dimensions the operator in (9.16) is acting 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 (7.7) is correct.

9.3.1 Matrix representations

When considering analytic maps there is another, and for numerical purposes, sometimes convenient way to look at the operators, namely through matrix representations. The size of these matrices is infinite but entries in the matrix decay exponentially fast with the indisize. Hence, within an exponentially small error one may safely do calculations using finite matrix truncations.

Furthermore, from bounds on the elements L_{mn} one calculates bounds on $\mathrm{tr}\left(\wedge^k \mathcal{L}\right)$ *and verifies that they fall off as $\Lambda^{-k^2/2}$, concluding that the $\mathcal{L}$ eigenvalues fall off exponentially for a general Axiom A 1-d map. In order to illustrate how this works, we work out a simple example.*

As in sect. 9.1 we start with a map with a single fixed point, but this time

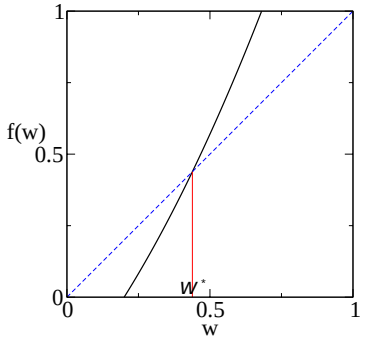


Figure 9.5: A nonlinear one-branch repeller with a single fixed point w^*.

with a nonlinear map f with a nonlinear inverse $F = f^{-1}$, $s = sgn(F')$

$$\mathcal{L} \circ \phi(z) = \int dx\ \delta(z - f(x))\, \phi(x) = s\ F'(z)\ \phi(F(z))\,.$$

Assume that F is a contraction of the unit disk, that is

$$|F(z)| < \theta < 1 \quad and \quad |F'(z)| < C < \infty \quad for \quad |z| < 1\,, \tag{9.17}$$

and expand ϕ in a polynomial basis by means of the Cauchy formula

$$\phi(z) = \sum_{n\geq 0} z^n \phi_n = \oint \frac{dw}{2\pi i}\ \frac{\phi(w)}{w - z}\,, \quad \phi_n = \oint \frac{dw}{2\pi i}\ \frac{\phi(w)}{w^{n+1}}$$

In this basis, $\mathcal{L}$ is a represented by the matrix

$$\mathcal{L} \circ \phi(w) = \sum_{m,n} w^m L_{mn} \phi_n\,, \quad L_{mn} = \oint \frac{dw}{2\pi i}\ \frac{s\ F'(w)(F(w))^n}{w^{m+1}}\,. \tag{9.18}$$

Taking the trace and summing we get:

$$\mathrm{tr}\ \mathcal{L} = \sum_{n\geq 0} L_{nn} = \oint \frac{dw}{2\pi i}\ \frac{s\ F'(w)}{w - F(w)}\,.$$

This integral has but one simple pole at the unique fix point $w^* = F(w^*) = f(w^*)$. Hence

$$\mathrm{tr}\ \mathcal{L} = \frac{s\ F'(w^*)}{1 - F'(w^*)} = \frac{1}{|f'(w^*) - 1|}\,.$$

9.6 on p. 195

We recognize this result as a generalization of the single piecewise-linear fixed-point example (9.2), $\phi_n = y^n$, and $\mathcal{L}$ is diagonal (no sum on repeated n here), $L_{nn} = 1/|\Lambda|\Lambda^{-n}$, so we have verified the heuristic trace formula for an expanding map with a single fixed point. The requirement that map be analytic is needed to substitute bound (9.17) into the contour integral (9.18) and obtain the inequality

$$|L_{mn}| \leq \sup_{|w|\leq 1} |F'(w)|\,|F(w)|^n \leq C\theta^n$$

which shows that finite $[N \times N]$ matrix truncations approximate the operator within an error exponentially small in N. It also follows that eigenvalues fall off as θ^n. In higher dimension similar considerations show that the entries in the matrix fall off as $1/\Lambda^{k^{1+1/d}}$, and eigenvalues as $1/\Lambda^{k^{1/d}}$.

9.4 Physics of eigenvalues and eigenfunctions

We appreciate by now that any serious attempt to look at 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 analyticity properties of dynamical zeta functions and spectral determinants, and thus substantiate the claim that these objects provide a powerful and well founded perturbation theory.

Quite often (see for instance chapter 6) the physical interest is concentrated in the leading eigenvalue, as it gives the escape rate from a repeller, or, when considering generalized transfer operators, it yields expressions for generating functions for observables. We recall (see chapter 5) that also the eigenfunction associated to the leading eigenvalue has a remarkable property: it provides the density of the invariant measure, with singular measures ruled out by the choice of the function space. Such a conclusion is coherent with a the validity of a generalized Perron-Frobenius theorem for the evolution operator. In the finite dimensional setting such theorem is formulated as follows:

- *let L_{nm} 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 is such that $0 < \rho_{ess} < \theta < 1$ *on some Banach space* $\mathcal{B}$ *and denote by* $\mathbf{P}$ *the projection corresponding to the part of the spectrum inside a disk of radius* θ*. We denote by* $\lambda_1, \lambda_2 \dots \lambda_M$ *the eigenvalues outside of this disk, ordered by the size of their absolute value (so that* $\lambda_1 = 1$*). Then we have the following decomposition*

$$\mathcal{L}\varphi = \sum_{i=1}^{M} \lambda_i \psi_i L_i \psi_i^* \varphi + \mathbf{P}\mathcal{L}\varphi \tag{9.19}$$

when L_i *are (finite) matrices in Jordan normal form (*$L_1 = 1$ *is a* 1×1 *matrix, as* λ_1 *is simple, due to Perron-Frobenius theorem), while* ψ_i *is a row vector whose elements are a basis on the eigenspace corresponding to* λ_i*, and* ψ_i^* *is a column vector of elements of* $\mathcal{B}^*$ *(the dual space, of linear functionals over* $\mathcal{B}$*) spanning the eigenspace of* $\mathcal{L}^*$ *corresponding to* λ_i*. For iterates of Perron-Frobenius operator (9.19) becomes*

$$\mathcal{L}^n\varphi = \sum_{i=1}^{M} \lambda_i^n \psi_i L_i^n \psi_i^* \varphi + \mathbf{P}\mathcal{L}^n\varphi \tag{9.20}$$

If we now consider expressions like

$$C(n)_{\xi,\varphi} = \int_{\mathcal{M}} dy\, \xi(y) \left(\mathcal{L}^n\varphi\right)(y) = \int_{\mathcal{M}} dw\, (\xi \circ f^n)(w)\varphi(w) \tag{9.21}$$

we have

$$C(n)_{\xi,\varphi} = \lambda_1^n \omega_1(\xi,\varphi) + \sum_{i=2}^{L} \lambda_i^n \omega(n)_i(\xi,\varphi) + \mathcal{O}(\theta^n) \tag{9.22}$$

where

$$\omega(n)_i(\xi,\varphi) = \int_{\mathcal{M}} dy\, \xi(y) \psi_i L_i^n \psi_i^* \varphi$$

In this way we see how eigenvalues beyond the leading one provide a twofold piece of information: they rule the convergence of expressions containing high powers of evolution operator to the leading order (the λ_1 *contribution). Moreover if* $\omega_1(\xi,\varphi) = 0$ *then (9.21) defines a correlation function: as each term in (9.22)*

9.7 on p. 195

vanishes exponentially in the $n \to \infty$ limit, the eigenvalues $\lambda_2, \ldots \lambda_M$ rule the exponential decay of correlations for our dynamical system. We observe that prefactors ω depend on the choice of functions, while the exponential decay rates (logarithms of λ_i) do not: the correlation spectrum is thus an universal *property of the dynamics (once we fix the overall functional space our Perron-Frobenius operator acts on).*

So let us come back the Bernoulli shift example (9.10), 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), we have that $\omega_1(\eta, \eta) = 0$, and from (9.22) we have that the asymptotic decay of correlation function is (unless also $\omega_1(\eta, \eta) = 0$)

$$C_{\eta,\eta}(n) \sim \exp(-n \log 2) \tag{9.23}$$

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

$$\eta(z) = \int_0^1 dw\, \eta(w) + \sum_{m=1}^{\infty} \frac{\eta^{(m-1)}(1) - \eta^{(m-1)}(0)}{m!} B_m(z)\,. \tag{9.24}$$

As we are considering zero–average functions, we have from (9.21), and the fact that Bernoulli polynomials are eigenvectors of the Perron-Frobenius operator

$$C_{\eta,\eta}(n) = \sum_{m=1}^{\infty} \frac{(2^{-m})^n (\eta^{(m)}(1) - \eta^{(m)}(0))}{m!} \int_0^1 dz\, \eta(z) B_m(z)\,.$$

The decomposition (9.24) is also useful to make us realize that the linear functionals ψ_i^* are quite 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

$$\Psi_i(w) \;=\; \frac{(-1)^{i-1}}{i!}\left(\delta^{(i-1)}(w-1) - \delta^{(i-1)}(w)\right) \tag{9.25}$$

when $i \geq 1$, *while* $\Psi_0(w) = 1$. *Such a representation is only meaningful when the function* ε *is analytic in* $w, w-1$ *neighborhoods.*

9.5 Why not just run it on a computer?

All of the insight gained in this chapter was nothing but an elegant way of thinking of $\mathcal{L}$ *as a matrix (and such a point of view will be further pursued in chapter 11). There are many textbook methods of approximating an operation* $\mathcal{L}$ *by sequences of finite matrix approximations* $\mathbf{L}$, *so why a new one?*

The simplest possible way of introducing a phase space discretization, fig. 9.6, is to partition the phase space $\mathcal{M}$ *with a non-overlapping collection of sets* $\mathcal{M}_\alpha$, $\alpha = 1, \ldots, N$, *and to consider densities that are locally constant on each* $\mathcal{M}_\alpha$*:*

$$\rho(x) \;=\; \sum_{\alpha=1}^{N} \wp_\alpha \frac{\chi_\alpha(x)}{m(\mathcal{A}_\alpha)}$$

where $\chi_\alpha(x)$ *is the characteristic function of the set* $\mathcal{A}_\alpha$. *Then the weights* $\wp_\alpha$ *are determined by the action of Perron-Frobenius operator*

$$\begin{aligned}\int_{\mathcal{M}} dz\, \chi_\beta(z)\rho(z) \;=\; \wp_\beta \quad &= \quad \int_{\mathcal{M}} dz\, \chi_\beta(z) \int_{\mathcal{M}} dw\, \delta(z - f(w))\, \rho(w) \\ &= \quad \sum_{\alpha=1}^{N} \wp_\alpha \frac{m(\mathcal{A}_\alpha \cap f^{-1}\mathcal{A}_\beta)}{m(\mathcal{A}_\alpha)}\end{aligned}$$

PCrewrite as in sect. 4.1 In this way

$$\mathbf{L}_{\alpha,\beta} \;=\; \frac{m(\mathcal{A}_\alpha \cap f^{-1}\mathcal{A}_\beta)}{m(\mathcal{A}_\alpha)} \tag{9.26}$$

is a matrix approximation to the Perron-Frobenius operator, and its 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.

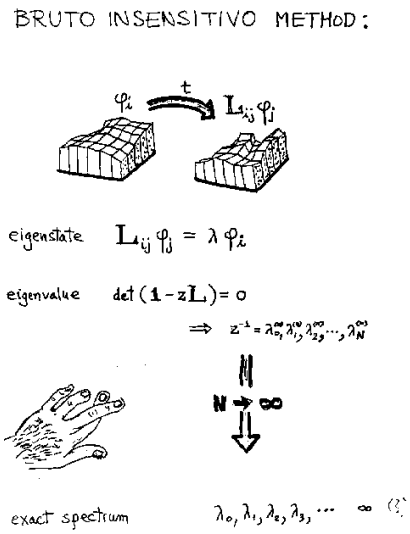


Figure 9.6: Phase space discretization approach to computing averages.

The problem with such general phase space discretization approaches is that they are blind; the grid knows not what parts of the phase space are more or less important, and with such methods one is often plagued by numerical artifacts such as spurious eigenvalues. In contrast, in this treatise we exploit the intrinsic topology of the flow to give us both an invariant partition of the phase space and invariant measure of the partition volumes, see fig. 1.8. We shall lean on the φ_α basis approach only insofar it helps us prove that the spectrum that we compute is indeed the correct one, and that finite periodic orbit truncations do converge.

Commentary

For a physicist Dricbee's monograph [] might be the most accessible introduction into main theories touched upon in this chapter.

Remark ***9.1*** Surveys of rigorous theory We recommend references listed in sect. **??** for an introduction into the mathematic literature on this subject. 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 illustrated in ref. [3]. The ergodic theory, as presented by Sinai [15] and others, tempts one to describe the densities that the evolution operator acts on in terms of either integrable or square integrable functions. As we have already seen, for our purposes, this space is not suitable. An introduction to ergodic theory is given by Sinai, Kornfeld and Fomin [16]; more advanced and more old fashioned presentations are Walters [17] and Denker, Grillenberger and Sigmund [18]; and a more formal Peterson [19].

PCgive credit to Prigople + ...(/)

Remark **9.2** Fredholm theory. Our brief summary of Fredholm theory is based on the exposition in ref. [4]. A technical introduction of the theory from an operatorial point of view is contained in ref. [5]. The theory has been generalized in ref. [6].

Remark **9.3** Bernoulli shift. For a more detailed discussion, consult chapter 15.1 or The extension of Fredholm theory to the case or Bernoulli shift on $C^{k+\alpha}$ (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 **9.4** Higher dimensions and generalized Fredholm theory. When extending Bernoulli shift to higher dimensions. Extensions of Fredholm theory [6], which avoid problems with multi-dimensional residue calculus, may be used: see ref. [9].

Remark **9.5** Hyperbolic dynamics. When dealing with hyperbolic systems one might try to reduce back to the expanding case by projecting the dynamics along the unstable directions. As mentioned in the text this might be technically quite involved, as usually such the unstable foliation is not characterized by very strong smoothness properties. For such an approach, see ref. [3].

Remark **9.6** Spectral determinants for smooth flows. The theorem on p. 169 applies also 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 suspension on a Poincaré section complemented by an analytic "ceiling" function (3.2) which 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 time of the next Poincaré section for a trajectory staring at x, we reproduce the flow spectral determinant (8.23). Proofs are getting too hard for the purposes of this chapter; details are discussed in ref.(?).

Remark **9.7** Examples. Examples of analytic hyperbolic maps are provided by small analytic perturbations of the cat map (where the Markov partitioning is non-trivial [10]), the 3-disk repeller, and the 2-d baker's map.

Remark **9.8** 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. [11, 12]). The reasons why one persists anyway in using the periodic orbit theory are partially aesthetic, and partially pragmatic. Explicit L_{mn} demands explicit choice of a basis and is thus non-invariant, in contrast to cycle expansions which utilize only the invariant information about the flow. In addition, we usually do not know how to construct L_{mn} for a realistic flow, such as the hyperbolic 3-disk game of pinball flow of sect. 1.3, whereas the periodic orbit formulas are general and straightforward to apply.

Remark **9.9** Perron-Frobenius theorem. A proof of the Perron-Frobenius theorem may be found in ref. [13]. For positive transfer operators such theorem has been generalized by Ruelle [14].

Remark **9.10** Fried estimates. The form of the fall-off of the coefficients in the $F(z)$ expansion, as $u^{n^{1+1/d}}$, is in agreement with the estimates of Fried [20] for the spectral determinants of d-dimensional expanding flows.

Remark **9.11** Axiom A systems. Proofs outlined in sect. 9.3 follow the thesis work of H.H. Rugh [9, 20, 21]. For mathematical introduction to the subject, consult the excellent review by V. Baladi [1]. Rigorous treatment is given in refs. [9, 20, 21]. It would take us too far to give and explain the definition of the Axiom A systems (see refs. [22, 23]). Axiom A implies, however, the existence of a Markov partition of the phase space from which the properties 2 and 3 assumed on p. 165 follow.

Remark **9.12** Exponential mixing speed of the Bernoulli shift. We see from (9.23) 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 **9.13** Left eigenfunctions. We shall never use explicit form of left eigenfunctions, corresponding to highly singular kernels like (9.25). Many details have been elaborated in a number of papers, like ref. [24], with a daring physical interpretation.

Remark **9.14** Ulam's idea. The approximation of Perron-Frobenius operator defined by (9.26) has been shown to reproduce correctly the spectrum for expanding maps, once finer and finer Markov partitions are used [25]. The subtle point of choosing a phase space partitioning for a "generic case" is discussed in ref. [26].

Résumé

A serious theory of cycle expansions requires a deeper understanding of their analyticity and convergence. If we restrict the considerations to those few ideal systems where symbolic dynamics and hyperbolicity can be controlled, it is possible to treat traces and determinants in a rigorous fashion, and beautiful rigorous results about 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 do they have a simple symbolic dynamics grammar. Importance of symbolic dynamics is sometime grossly unappreciated; the crucial ingredient for nice analyticity properties of zeta functions is existence of finite grammar (coupled with uniform hyperbolicity). The dynamical systems that we are *really* interested in - for example, smooth bound Hamiltonian potentials - are presumably never really chaotic, and the central question remains: how to attack the problem in systematic and controllable fashion?

References

[9.1] V. Baladi, *A brief introduction to dynamical zeta functions*, in: DMV-Seminar 27, *Classical Nonintegrability, Quantum Chaos*, A. Knauf and Ya.G. Sinai (eds), (Birkhuser,1997).

[9.2] 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.*

[9.3] M. Viana, *Stochastic dynamics of deterministic systems*, (Col. Bras. de Matemática, Rio de Janeiro,1997)

[9.4] A.N. Kolmogorov and S.V. Fomin, *Elements of the theory of functions and functional analysis* (Dover,1999).

[9.5] R.G. Douglas, *Banach algebra techniques in operator theory* (Springer, New York,1998).

[9.6] A. Grothendieck, *La théorie de Fredholm*, *Bull. Soc. Math. France* **84**, 319 (1956).

[9.7] D. Ruelle, *Inst. Hautes Études Sci. Publ. Math.* **72**, 175-193 (1990).

[9.8] V. Baladi, *Dynamical zeta functions*, Proceedings of the NATO ASI *Real and Complex Dynamical Systems* (1993), B. Branner and P. Hjorth, eds. (Kluwer Academic Publishers, Dordrecht, 1995)

[9.9] D. Ruelle, *Inv. Math.* **34**, 231-242 (1976).

[9.10] R.L. Devaney, *An Introduction to Chaotic Dynamical Systems* (Addison-Wesley, Reading MA, 1987).

[9.11] F. Christiansen, P. Cvitanović and H.H. Rugh, *J. Phys* **A 23**, L713 (1990).

[9.12] D. Alonso, D. MacKernan, P. Gaspard and G. Nicolis, *Phys. Rev.* **E54**, 2474 (1996).

[9.13] P. Walters, *An introduction to ergodic theory.* (Springer, New York 1982).

[9.14] D. Ruelle, *Commun. Math. Phys.* **9**, 267 (1968).

[9.15] Ya.G. Sinai, *Topics in ergodic theory.* (Princeton Univ. Press, Princeton 1994).

[9.16] I. Kornfeld, S. Fomin and Ya. Sinai, *Ergodic Theory* (Springer, 1982).

[9.17] P. Walters, *An introduction to ergodic theory*, Springer Graduate Texts in Math. Vol **79** (Springer, New York, 1982).

[9.18] M. Denker, C. Grillenberger and K. Sigmund, *Ergodic theory on compact spaces*, (Springer Lecture Notes in Math. **470**, 1975).

[9.19] K. Peterson, *Ergodic theory* (Cambridge Univ. Press, Cambridge 1983).

[9.20] D. Fried, *Ann. Scient. Éc. Norm. Sup.* **19**, 491 (1986).

[9.21] H.H. Rugh, *Nonlinearity* **5**, 1237 (1992).

[9.22] S. Smale, *Bull. Amer. Math. Soc.* **73**, 747 (1967).

[9.23] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, *Springer Lect. Notes in Math.* **470**, 1975.

[9.24] H.H. Hasegawa and W.C. Saphir, *Phys. Rev.* **A46**, 7401 (1992).

[9.25] G. Froyland, *Commun. Math. Phys.* **189**, 237 (1997)

[9.26] G. Froyland, *Extracting dynamical behaviour via markov models*, in A. Mees (ed.) *Nonlinear dynamics and statistics: Proceedings Newton Institute, Cambridge 1998* (Birkhauser,2000).

[9.27] V. Baladi, A. Kitaev, D. Ruelle, and S. Semmes, *"Sharp determinants and kneading operators for holomorphic maps"*, IHES preprint (1995).

[9.28] A. Zygmund, *Trigonometric series* (Cambridge Univ. Press, Cambridge 1959).

Exercises

9.1 What space does $\mathcal{L}$ act on? Show that (9.2) is a complete basis on the space of analytic functions on a disk (and thus that we found the *complete* set of eigenvalues).

9.2 What space does $\mathcal{L}$ act on? What can be said about the spectrum of (9.1) on $L^1[0,1]$? Compare the result with fig. 9.1.

9.3 Euler formula. Derive the Euler formula (9.4)

$$\begin{aligned}\prod_{k=0}^{\infty}(1+tu^k) &= 1+\frac{t}{1-u}+\frac{t^2u}{(1-u)(1-u^2)}+\frac{t^3u^3}{(1-u)(1-u^2)(1-u^3)}\cdots \\ &= \sum_{k=0}^{\infty} t^k \frac{u^{\frac{k(k-1)}{2}}}{(1-u)\cdots(1-u^k)}\,, \qquad |u|<1. \end{aligned} \tag{9.27}$$

9.4 2-d product expansion**. We conjecture that the expansion corresponding to (9.27) is in this case

$$\begin{aligned}\prod_{k=0}^{\infty}(1+tu^k)^{k+1} &= \sum_{k=0}^{\infty}\frac{F_k(u)}{(1-u)^2(1-u^2)^2\cdots(1-u^k)^2}t^k \\ &= 1+\frac{1}{(1-u)^2}t+\frac{2u}{(1-u)^2(1-u^2)^2}t^2 \\ &\quad +\frac{u^2(1+4u+u^2)}{(1-u)^2(1-u^2)^2(1-u^3)^2}t^3+\cdots \end{aligned} \tag{9.28}$$

$F_k(u)$ 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 9.3).

9.5 Bernoulli shift on L spaces. Check that the family (9.12) belongs to $L^1([0,1])$. What can be said about the essential spectral radius on $L^2([0,1])$? A useful reference is [28].

9.6 Cauchy integrals. Rework all complex analysis steps used in the Bernoulli shift example on analytic functions on a disk.

9.7 Escape rate. Consider the escape rate from a strange repeller: find a choice of trial functions ξ and φ such that (9.21) gives the fraction on particles surviving after n iterations, if their initial density distribution is $\rho_0(x)$. Discuss the behavior of such an expression in the long time limit.

Chapter 10

Qualitative dynamics

> The classification of the constituents of a chaos, nothing less is here essayed.
> Herman Melville, *Moby Dick*, chapter 32

In chapters 7 and 8 we established that spectra of evolution operators can be extracted from periodic orbit sums:

$$\sum(\text{eigenvalues}) \;=\; \sum(\text{periodic orbits}) \;\;.$$

In order to apply this theory we need to know what periodic orbits can exist.

In this chapter and the next we learn how to *name* and *count* periodic 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 showing that the qualitative dynamics of stretching and mixing flows enables us to partition the phase 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. We turn this topological dynamics into a multiplicative operation by means of transition matrices/Markov graphs.

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 the next are good for you. Read them.

10.1 Temporal ordering: Itineraries

(R. Mainieri and P. Cvitanović)

What can a flow do to the phase 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 as long as it encloses x_0, and call it the neighborhood $\mathcal{M}_0$. 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 phase 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*. Understanding symbolic dynamics is a prerequisite to developing a theory of chaotic dynamic systems. We offer a summary of the basic notions and definitions of symbolic dynamics in sect. 10.2. As the subject can get quite technical, you might want to skip this section on first reading, but check there whenever you run into obscure symbolic dynamics jargon.

We start by cutting up the phase 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 phase 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 phase 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 fig. 10.1, the phase 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 a possible itinerary for the trajectory of a point x would be $0 \mapsto 2 \mapsto 1 \mapsto 0 \mapsto 1 \mapsto 2 \mapsto \cdots$.

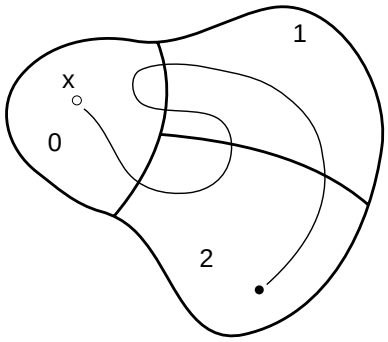


Figure 10.1: A trajectory with itinerary 021012.

An interesting partition should be dynamically connected, that is 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 *metrically indecomposable*.

The allowed transitions between the regions of a partition are encoded in the $[N\times N]$-dimensional *transition matrix* whose elements take values

$$T_{ij} = \begin{cases} 1 & \text{if a transition region } \mathcal{M}_j \to \text{region } \mathcal{M}_i \text{ is possible} \\ 0 & \text{otherwise}. \end{cases} \tag{10.2}$$

An example is the complete N-ary dynamics for which all transition matrix entries equal unity (one can reach any region to 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 (10.14) and the 1-step memory sparse matrix (10.27), are peppered throughout the text. 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.

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.

Knowing that some 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*.

Such considerations motivate the notion of a *Markov partition*, a partition for which no memory of preceeding 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,

$$\text{either} \quad \mathcal{M}_j \cap f(\mathcal{M}_i) = \emptyset \quad \text{or} \quad \mathcal{M}_j \subset f(\mathcal{M}_i)\,. \tag{10.4}$$

An example is the 1-dimensional expanding mapping sketched in fig. 10.6, and more examples are worked out in sect. 18.2.

fast track:
sect. 10.3, p. 204

10.2 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.

Definitions.

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

$$x_{n+1} = f(x_n)\,, \tag{10.5}$$

then the itinerary is given by the symbol sequence

$$s_n = s \qquad \text{if} \qquad x_n \in \mathcal{M}_s\,. \tag{10.6}$$

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

$$S(x_0) = (s_k)_{k\in\mathbb{Z}} = S^-.S^+ = \cdots s_{-2} s_{-1} s_0 . s_1 s_2 s_3 \cdots\,. \tag{10.7}$$

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*

$$\mathcal{A}^{\mathbb{Z}} = \{(s_k)_{k\in\mathbb{Z}} : s_k \in \mathcal{A} \text{ for all k } \in \mathbb{Z}\}\,. \tag{10.8}$$

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 (10.6), 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 σ,

$$\sigma(\cdots s_{-2}s_{-1}s_0.s_1s_2s_3\cdots) = \cdots s_{-2}s_{-1}s_0s_1.s_2s_3\cdots\,, \tag{10.9}$$

demoting the current partition label s_1 from the future S^+ to the "has been" itinerary S^-. The inverse shift σ^{-1} shifts the entire itinerary one step to the right.

A finite sequence $b = s_k s_{k+1}\cdots s_{k+n_b-1}$ of symbols from $\mathcal{A}$ is called a *block* of length n_b. A phase 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^∞. We shall refer to the set of periodic points that belong to a given periodic orbit as a *cycle*

$$p = \overline{s_1 s_2\cdots s_{n_p}} = \{x_{s_1 s_2\cdots s_{n_p}}, x_{s_2\cdots s_{n_p}s_1}, \cdots, x_{s_{n_p}s_1\cdots s_{n_p-1}}\}\,. \tag{10.10}$$

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 labelled 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).

A partition is called *generating* if every infinite symbol sequence corresponds to a distinct point in the phase 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* (Σ, σ), where the *subshift*

$$\Sigma = \{(s_k)_{k\in\mathbb{Z}}\}\,, \tag{10.11}$$

is the set of all *admissible* infinite itineraries, and $\sigma : \Sigma \to \Sigma$ is the shift operator (10.9). The designation "subshift" comes form the fact that $\Sigma \subset \mathcal{A}^{\mathbb{Z}}$ is the subset of the full shift (10.8). One of our principal tasks in developing symbolic dynamics of dynamical systems that occur in nature will be to determine Σ, 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 $\cdots s_{-2}s_{-1}s_0.s_1s_2s_3\cdots$ to each distinct trajectory, but there might exist full shift symbol sequences (10.8) 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 will be explained in sect. 10.8.

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,

$$\mathcal{G} = \{b_1, b_2, \cdots b_k\}\ , \tag{10.12}$$

where a *pruning block* b is a sequence of symbols $b = s_1s_2\cdots s_{n_b}$, $s \in \mathcal{A}$, 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.

(a) $T = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ (b)

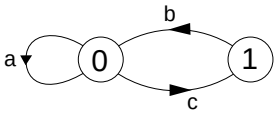


Figure 10.2: (a) The transition matrix for a simple subshift on two-state partition $\mathcal{A} = \{0, 1\}$, with grammar $\mathcal{G}$ given by a single pruning block $b = 11$ (consecutive repeat of symbol 1 is inadmissible): the state $\mathcal{M}_0$ maps both onto $\mathcal{M}_0$ and $\mathcal{M}_1$, but the state $\mathcal{M}_1$ maps only onto $\mathcal{M}_0$. (b) The corresponding finite 2-node, 3-links Markov graph, with nodes coding the symbols. All admissible itineraries are generated as walks on this finite Markov graph.

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 (Σ, σ) for which all admissible itineraries are generated by a finite transition matrix

$$\Sigma = \left\{ (s_k)_{k\in\mathbb{Z}} \,:\, T_{s_k s_{k+1}} = 1 \quad \text{for all } k \right\} \tag{10.13}$$

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.

A Markov graph describes compactly the ways in which the phase-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 relabelling 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, that is graphs for which there is a path from any node to any other node. irreducible!graph strongly connected graph graph!irreducible

The simplest example is given in fig. 10.2. We shall study such graphs in more detail in sect. 10.8.

10.3 3-disk symbolic dynamics

The key symbolic dynamics concepts are easily illustrated by a game of pinball. Consider the motion of a free point particle in a plane with N elastically reflecting convex disks. After a collision with a disk a particle either continues to another disk or escapes, and any trajectory can be labelled by the disk sequence. For example, if we label the three disks by 1, 2 and 3, the two trajectories in fig. 1.2 have itineraries _3123_, _312132_ respectively. The 3-disk prime cycles given in figs. 1.4 and 10.4 are further examples of such itineraries.

1.1 on p. 32

At each bounce a pencil of initially nearby trajectories defocuses, and in order to aim at a desired longer and longer itinerary of bounces the initial point $x_0 = (p_0, q_0)$ has to be specified with a larger and larger precision. 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 $x_0 = (p_0, q_0)$ 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 $\sin\theta \to -\sin\theta$). $\mathcal{M}_{s_1.s_2}$ is a "rectangle" of nearby trajectories which have arrived from the disk s_1 and are heading for the disk s_2.

We see that a finite length trajectory is not uniquely specified by its finite itinerary, but an isolated unstable cycle (consisting of infinitely many repetitions of a prime building block) is, and so is a trajectory with a bi-infinite itinerary $S^{\text{-}}.S^{+} = \cdots s_{-2}s_{-1}s_0.s_1s_2s_3\cdots$. For hyperbolic flows the intersection of the future and past itineraries uniquely specifies a 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 pencil of trajectories with a given itinerary becomes sharper and sharper as the number of specified symbols is increased.

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 (for the definition, see (10.13)), with the transition matrix (10.2) given by

$$T = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} . \tag{10.14}$$

For convex disks the separation between nearby trajectories increases at every reflection, implying that the stability matrix has an expanding eigenvalue. By

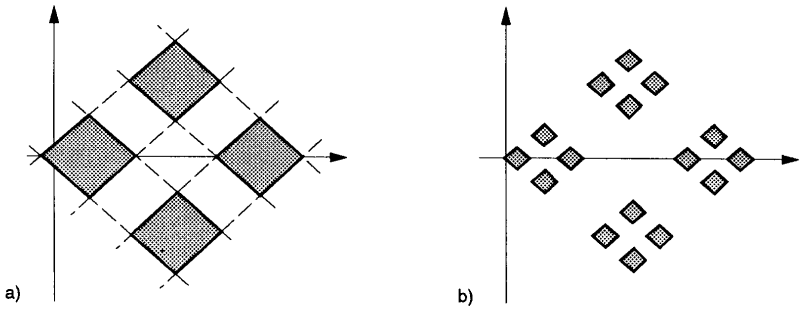


Figure 10.3: The Poincaré section of the phase space for the binary labelled pinball, see also fig. 10.4(b). Indicated are the fixed points $\overline{0}$, $\overline{1}$ and the 2-cycle periodic points $\overline{01}$, $\overline{10}$, together with strips which survive 1, 2, ... bounces. Iteration corresponds to the decimal point shift; for example, all points in the rectangle $[01.01]$ map into the rectangle $[010.1]$ in one iteration.
PC: do this figure right, in terms of strips!

the Liouville phase-space volume conservation (4.39), the other transverse eigenvalue is contracting. This example shows that finite Markov partitions can be constructed for *hyperbolic* dynamical systems which are expanding in some directions, contracting in others.

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. For the time being we assume that the disks are sufficiently separated that there is no additional pruning beyond the prohibition of self-bounces.

fast track:
sect. 10.5, p. 210

10.3.1 A brief detour; nonuniqueness, symmetries, tilings

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. As we shall see in chapter 17, rewards of this desymmetrization will be handsome.

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 relabelling of the disks. Further examples of such symmetries are shown in fig. 1.4. We note that 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, fig. 1.3:

10.1 on p. 233

0: the pinball returns to the disk it came from
1: the pinball continues to the third disk.

10.2 on p. 233

This *binary* symbolic dynamics has one immediate advantage over the ternary one; the prohibition of self-bounces is automatic. If the disks are sufficiently far apart there are no further restrictions on symbols, the symbolic dynamics is complete, and *all* binary sequences are admissible itineraries. As this type of symbolic dynamics pops up frequently, we list the shortest binary prime cycles in table 10.1.

10.3 on p. 233

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 fig. 10.4b. A global 3-disk trajectory maps into its fundamental domain mirror trajectory by replacing every crossing of a symmetry axis by a reflection. Depending on the symmetry of the global trajectory, a repeating binary symbols block corresponds either to the full periodic orbit or to an irreducible segment (examples are shown in fig. 10.4 and table 10.2). An irreducible segment corresponds to a periodic orbit in the fundamental domain. Table 10.2 lists some of the shortest binary periodic orbits, together with the corresponding full 3-disk symbol sequences and orbit symmetries. For a number of reasons that will be elucidated in chapter 17, 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.

10.4 on p. 234

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 the symbolic dynamics. This is brought out more clearly by the Smale horseshoe visualization of "stretch & fold" flows to which we turn now.

10.4 Spatial ordering of "stretch & fold" flows

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. So

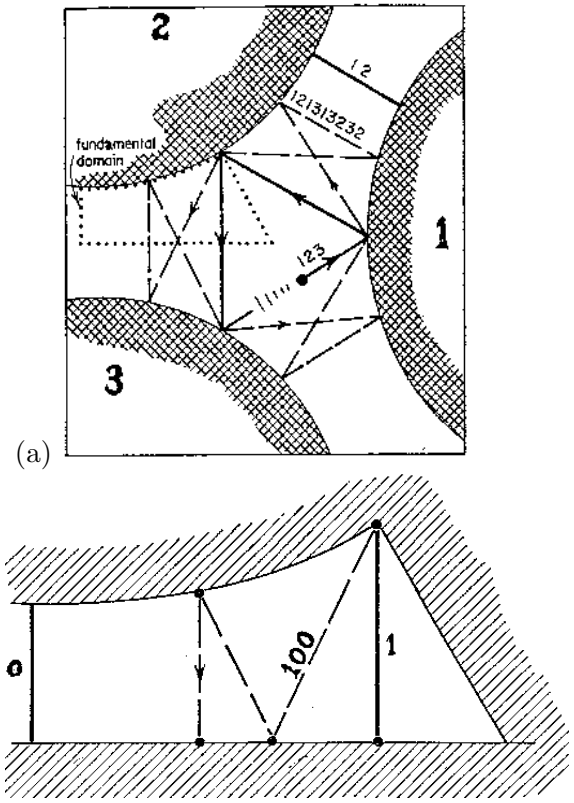


Figure 10.4: The 3-disk game of pinball with the disk radius : center separation ratio a:R = 1:2.5. (a) The three disks, with $\overline{12}$, $\overline{123}$ and $\overline{121232313}$ cycles indicated. (b) The fundamental domain, that is the small 1/6th wedge indicated in (a), consisting of a section of a disk, two segments of symmetry axes acting as straight mirror walls, and an escape gap. The above cycles restricted to the fundamental domain are now the two fixed points $\overline{0}$, $\overline{1}$, and the $\overline{100}$ cycle.

n_p	p
1	0
	1
2	01
3	001
	011
4	0001
	0011
	0111
5	00001
	00011
	00101
	00111
	01011
	01111
6	000001
	000011
	000101
	000111
	001011
	001101
	001111
	010111
	011111
7	0000001
	0000011
	0000101

n_p	p
7	0001001
	0000111
	0001011
	0001101
	0010011
	0010101
	0001111
	0010111
	0011011
	0011101
	0101011
	0011111
	0101111
	0110111
	0111111
8	00000001
	00000011
	00000101
	00001001
	00000111
	00001011
	00001101
	00010011
	00010101
	00011001
	00100101

n_p	p
8	00001111
	00010111
	00011011
	00011101
	00100111
	00101011
	00101101
	00110101
	00011111
	00101111
	00110111
	00111011
	00111101
	01010111
	01011011
	00111111
	01011111
	01101111
	01111111
9	000000001
	000000011
	000000101
	000001001
	000010001
	000000111
	000001011

n_p	p
9	000001101
	000010011
	000010101
	000011001
	000100011
	000100101
	000101001
	000001111
	000010111
	000011011
	000011101
	000100111
	000101011
	000101101
	000110011
	000110101
	000111001
	001001011
	001001101
	001010011
	001010101
	000011111
	000101111
	000110111
	000111011
	000111101

n_p	p
9	001001111
	001010111
	001011011
	001011101
	001100111
	001101011
	001101101
	001110101
	010101011
	000111111
	001011111
	001101111
	001110111
	001111011
	001111101
	010101111
	010110111
	010111011
	001111111
	010111111
	011011111
	011101111
	011111111

Table 10.1: Prime cycles for the binary symbolic dynamics up to length 9.

a typical dynamical system that we care about is *bounded*. If the price for change is high - for example, we try to stir up some tar, and observe it come to dead stop the moment we cease our labors - the dynamics tends to settle into a simple limiting state. However, as the resistence to change decreases - the tar is heated up and we are more vigorous in our stirring - the dynamics becomes unstable. We have already quantified this instability in sect. 4.1 - for now suffice it to say that a flow is *locally unstable* if nearby trajectories separate exponentially with time.

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 fig. 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 fig. 10.16). 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 system given below in (2.12).

10.6 on p. 234

At this juncture the reader can chose either of the paths illustrating the concepts introduced above, or follow both: a shortcut via unimodal mappings of the interval, sect. 10.5, or more demanding path, via the Smale horseshoes of

$\tilde{p}$	p	$\mathbf{g}_{\tilde{p}}$
0	1 2	σ_{12}
1	1 2 3	C_3
01	12 13	σ_{23}
001	121 232 313	C_3
011	121 323	σ_{13}
0001	1212 1313	σ_{23}
0011	1212 3131 2323	C_3^2
0111	1213 2123	σ_{12}
00001	12121 23232 31313	C_3
00011	12121 32323	σ_{13}
00101	12123 21213	σ_{12}
00111	12123	e
01011	12131 23212 31323	C_3
01111	12132 13123	σ_{23}

$\tilde{p}$	p	$\mathbf{g}_{\tilde{p}}$
000001	121212 131313	σ_{23}
000011	121212 313131 232323	C_3^2
000101	121213	e
000111	121213 212123	σ_{12}
001011	121232 131323	σ_{23}
001101	121231 323213	σ_{13}
001111	121231 232312 313123	C_3
010111	121312 313231 232123	C_3^2
011111	121321 323123	σ_{13}
0000001	1212121 2323232 3131313	C_3
0000011	1212121 3232323	σ_{13}
0000101	1212123 2121213	σ_{12}
0000111	1212123	e
...	...	...

Table 10.2: C_{3v} correspondence between the binary labelled fundamental domain prime cycles $\tilde{p}$ and the full 3-disk ternary labelled 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. 17.2.2. Breaks in the ternary sequences mark repeats of the irreducible segment. The degeneracy of p cycle is $m_p = 6n_{\tilde{p}}/n_p$. The shortest pair of the fundamental domain cycles related by time symmetry are the 6-cycles $\overline{001011}$ and $\overline{001101}$.

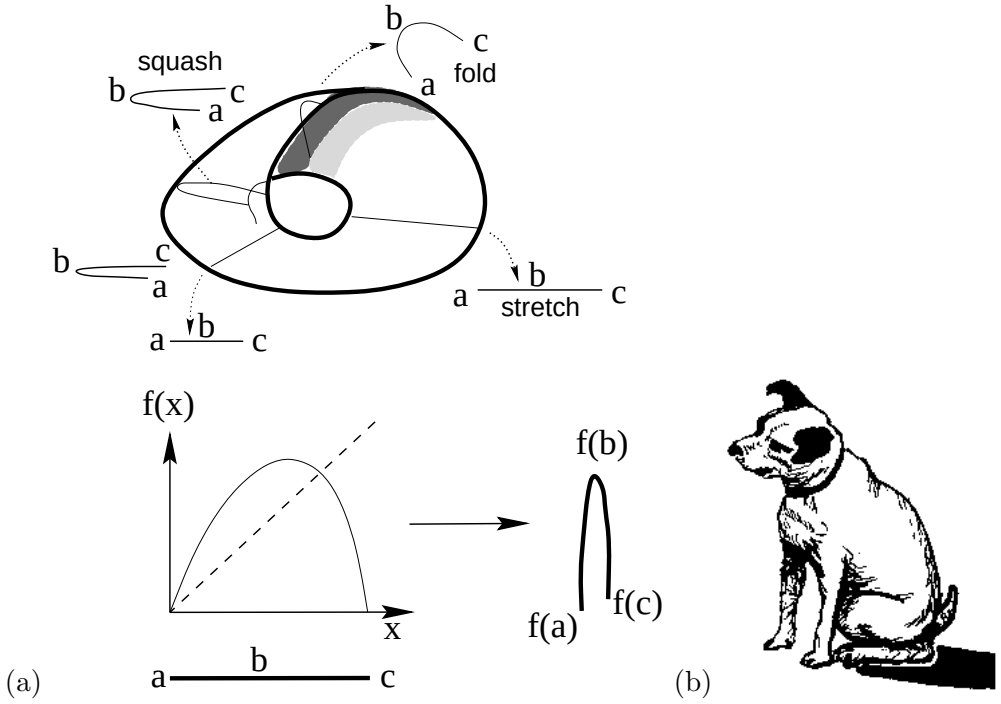


Figure 10.5: (a) A recurrent flow that stretches and folds. (b) The "stretch & fold" return map on the Poincaré section.

sects. 10.6 and 10.7. Unimodal maps are easier, but physically less motivated. The Smale horseshoes are the high road, more complicated, but the right tool to describe the 3-disk dynamics, and begin analysis of general dynamical systems. It is up to you - to get quickly to the next chapter, unimodal maps will suffice.

in depth:
sect. 10.6, p. 215

10.5 Unimodal map symbolic dynamics

Our next task is to relate the spatial ordering of phase-space points to their temporal itineraries. 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 symbols, let us state the main result at the outset: the admissibility criterion stated in sect. 10.5.2 eliminates *all* itineraries that cannot occur for a given unimodal map.

Suppose that the compression of the folded interval in fig. 10.5 is so fierce that we can neglect the thickness of the attractor. For example, the Rössler flow (2.12) 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. 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 new refolded interval. A *unimodal* map $f(x)$ is a 1-d function $\mathbb{R} \to \mathbb{R}$ defined on an interval $\mathcal{M}$ 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. The name is uninspiring - it refers to any one-humped map of interval into itself.

The simplest examples of unimodal maps are the *complete tent map* fig. 10.6(a),

$$f(\gamma) = 1 - 2|\gamma - 1/2|\,, \tag{10.15}$$

and the quadratic map (sometimes also called the logistic map)

$$x_{t+1} = 1 - ax_t^2\,, \tag{10.16}$$

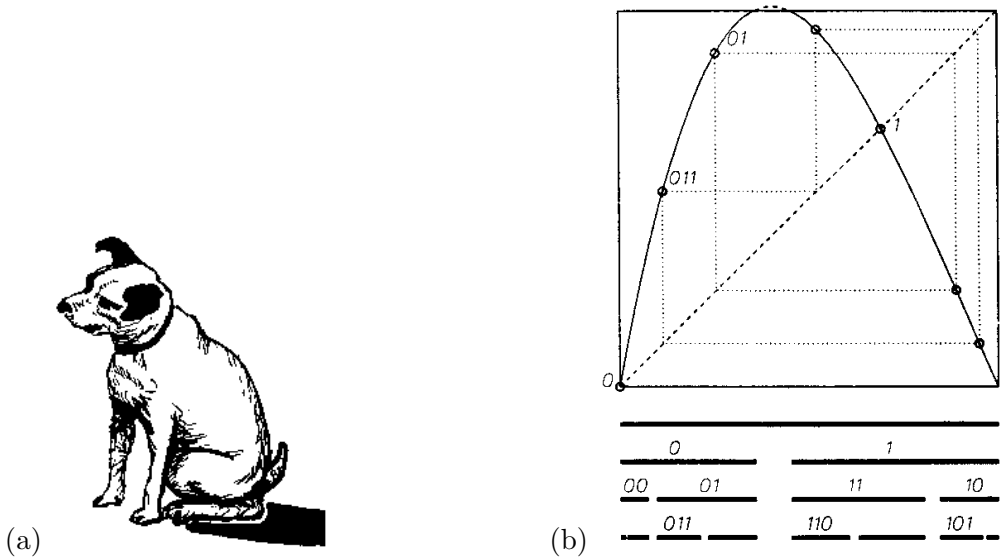


Figure 10.6: (a) The complete tent map together with intervals that follow the indicated itinerary for n steps. (b) A unimodal repeller with the remaining intervals after 1, 2 and 3 iterations. Intervals marked $s_1 s_2 \cdots s_n$ are unions of all points that do not escape in n iterations, and follow 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 $\overline{011}$. (need correct eq. ref.)

with the one critical point at $x_c = 0$. Another example is the repelling unimodal map of fig. 10.6b. We refer to (10.15) as the "complete" tent map because its symbolic dynamics is a complete binary dynamics.

Such dynamical systems are irreversible (the inverse of f is double-valued), but, as we shall argue in sect. 10.6.1, they may nevertheless serve as effective descriptions of hyperbolic flows. For the unimodal maps of fig. 10.6 a Markov partition of the unit interval $\mathcal{M}$ is given by the two intervals $\{\mathcal{M}_0, \mathcal{M}_1\}$. The 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). 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 $s = C$.

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 (10.6) for a given trajectory,

$$s_n = \begin{cases} 1 & \text{if } x_n > x_c \\ 0 & \text{if } x_n < x_c \end{cases} . \tag{10.17}$$

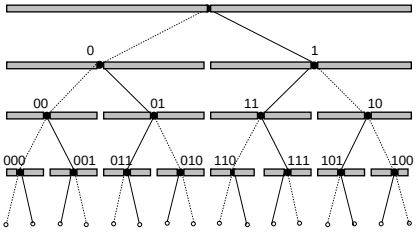


Figure 10.7: Alternating binary tree relates the itinerary labelling of the unimodal map fig. 10.6 intervals to their spatial ordering. Dotted line stands for 0, full line for 1; the binary sub-tree whose root is a full line (symbol 1) reverses the orientation, due to the orientation reversing fold in figs. 10.6 and 10.5.

We shall refer to $S^+(x_0) = .s_1 s_2 s_3 \cdots$ as the future itinerary. Our next task is answer the reverse problem: given an itinerary, what is the corresponding *spatial* ordering of points that belong to a given trajectory?

10.5.1 Spatial ordering for unimodal mappings

The tent map (10.15) consists of two straight segments joined at $x = 1/2$. The symbol s_n defined in (10.17) 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 problems of combinatorics and symbolic dynamics: 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. 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:

$$\begin{aligned} w_{n+1} &= \begin{cases} w_n & \text{if } s_n = 0 \\ 1 - w_n & \text{if } s_n = 1 \end{cases}, \qquad w_1 = s_1 \\ \gamma(S^+) &= 0.w_1 w_2 w_3 \ldots = \sum_{n=1}^{\infty} w_n / 2^n . \end{aligned} \tag{10.18}$$

10.5 on p. 234 This follows by inspection from the binary tree of fig. 10.7. For example, γ 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 nothing more than digits in the binary expansion of the starting point γ for the complete tent map (10.15). 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 γ for which the trajectory of the "canonical" unimodal map (10.15) 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 $\overline{\gamma} > \gamma$.

10.5.2 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 (10.16) 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 labelled Cantor set, the $n \to \infty$ limit of the nth level covering intervals sketched in fig. 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*

$$\kappa = \gamma(K) = \gamma(S^+(x_c)). \tag{10.19}$$

A map with the same kneading sequence K as $f(x)$, such as the dike map fig. 10.8, is obtained by slicing off all $\gamma\left(S^+(x_0)\right) > \kappa$,

$$f(\gamma) = \begin{cases} f_0(\gamma) = 2\gamma & \gamma \in I_0 = [0, \kappa/2) \\ f_c(\gamma) = \kappa & \gamma \in I_c = [\kappa/2, 1 - \kappa/2] \\ f_1(\gamma) = 2(1 - \gamma) & \gamma \in I_1 = [1 - \kappa/2, 1] \end{cases} . \tag{10.20}$$

The dike map is the complete tent map fig. 10.6(a) with the top sliced off. It is convenient for coding the symbolic dynamics, as those γ values that survive the pruning are the same as for the complete tent map fig. 10.6(a), and are easily converted into admissible itineraries by (10.18).

Figure 10.8: The "dike" map obtained by slicing of a top portion of the tent map fig. 10.6a. 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 fig. 10.6a orbit.

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

$$\hat{\gamma}(S^+) = \sup_m \gamma(\sigma^m(S^+)) \tag{10.21}$$

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 γ of any shifted sequence of S^+ falls into this interval.

Criterion of admissibility: *Let κ be the kneading value of the critical point, and $\hat{\gamma}(S^+)$ be the maximal value of the orbit S^+. Then the orbit S^+ is admissible if and only if $\hat{\gamma}(S^+) \leq \kappa$.*

While a unimodal map may depend on many arbitrarily chosen parameters, its dynamics determines the unique kneading value κ. We shall call κ the *topological parameter* of the map. Unlike the parameters of the original dynamical system, the topological parameter has no reason to be either smooth or continuous. The jumps in κ as a function of the map parameter such as a in (10.16) correspond to inadmissible values of the topological parameter. Each jump in κ corresponds to a stability window associated with a stable cycle of a smooth unimodal map. For the quadratic map (10.16) κ increases monotonically with the parameter a, but for a general unimodal map monotonicity need not be the case.

For further details of unimodal dynamics, the reader is referred to appendix E.1. As we shall see in sect. 10.7, 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.

Armed with one example of pruning, the impatient reader might prefer to skip the 2-dimensional examples and jump from here directly to the topological dynamics sect. 10.8.

fast track:
sect. 10.8, p. 222

10.6 Spatial ordering: Symbol square

> **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, that is, all of the orbits of M.*
>
> Stephen Smale, *Differentiable Dynamical Systems*

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.

So far we have rules that, given a phase 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".

10.6.1 Horseshoes

In fig. 10.5 we gave an example of a locally unstable but globally bounded flow 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. We shall refer to such flow-induced mappings from a Poincaré section to itself with at most 2^n transverse intersections at the nth iteration as the *once-folding* maps.

3.4 on p. 70

As an example of a flow for which the iterates of an initial region intersect as claimed above, consider the 2-dimensional *Hénon map*

$$\begin{aligned} x_{n+1} &= 1 - ax_n^2 + by_n \\ y_{n+1} &= x_n \,. \end{aligned} \tag{10.22}$$

The Hénon map models qualitatively the Poincaré section return map of fig. 10.5. For $b = 0$ the Hénon map reduces to the parabola (10.16), and, as we shall see here and in sects. 3.3 and 12.4.1, for $b \neq 0$ it is kind of a fattened parabola; it takes a rectangular initial area and returns it bent as a horseshoe.

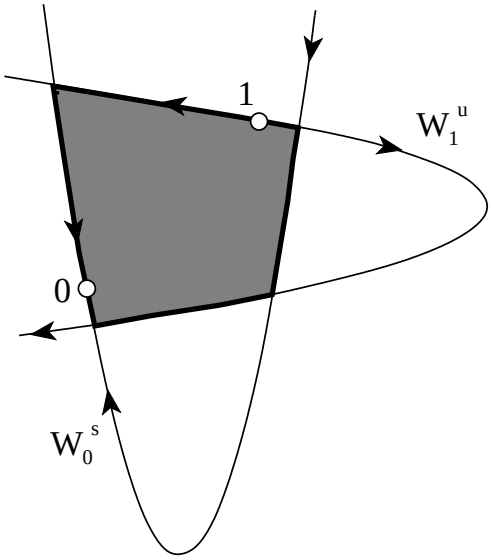
For definitiveness, fix the parameter values to $a = 6$, $b = 0.9$. The map is quadratic, so it has 2 fixed points $x_0 = f(x_0)$, $x_1 = f(x_1)$ indicated in fig. 10.9a. For the parameter values at hand, they are both unstable. If you start with a small ball of initial points centered around x_1, and 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,

$$\begin{aligned} x_{n-1} &= x_n \\ y_{n-1} &= -\frac{1}{b}(1 - ay_n^2 - x_n) \,, \end{aligned} \tag{10.23}$$

traces out the line W_0^s. W_0^s is the stable manifold of x_0, and W_1^u is the unstable manifold of x_1 fixed point (see sect. 4.8 - for now just think of them as curves going through the fixed points). Their intersection delineates the crosshatched region $\mathcal{M}_{\cdot}$. It is easily checked that any point outside W_1^u segments of the $\mathcal{M}_{\cdot}$ border escapes to infinity forward in time, while any point outside W_0^s border segments escapes to infinity backwards in time. That makes $\mathcal{M}_{\cdot}$ a good choice of the initial region; all orbits that stay confined for all times must be within $\mathcal{M}_{\cdot}$.

Iterated one step forward, the region $\mathcal{M}_{\cdot}$ is stretched and folded into a horseshoe as in fig. 10.9b. 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 weak compression and strong stretching. Denote the forward intersections $f(\mathcal{M}_{\cdot}) \cap \mathcal{M}_{\cdot}$ by $\mathcal{M}_{s\cdot}$, with $s \in \{0, 1\}$, fig. 10.9b. 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, fig. 10.9c. As stability and instability are interchanged under 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



(a) (b) (c)

Figure 10.9: (a) The Hénon map for $a = 6$, $b = .9$. Indicated are the fixed points $\overline{0}$, $\overline{1}$, and the segments of the W_0^s stable manifold, W_1^u unstable manifold that enclose the initial (crosshatched) region $\mathcal{M}_{.}$. (b) The forward horseshoe $f(\mathcal{M}_{.})$. (c) The backward horseshoe $f^{-1}(\mathcal{M}_{.})$. Iteration yields a complete Smale horseshoe, with every forward fold intersecting every backward fold.

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. 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 phase space into two regions labelled 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 mth level 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

$$\mathcal{M}_{0.} = f(\mathcal{M}_{.0})\,, \qquad \mathcal{M}_{1.} = f(\mathcal{M}_{.1})\,.$$

Ω, the non–wandering set (2.2) of $\mathcal{M}_{.}$, is the union of all the non-wandering points given by the intersections

$$\Omega = \left\{ x : x \in \lim_{m,n\to\infty} f^m(\mathcal{M}_{.}) \bigcap f^{-n}(\mathcal{M}_{.}) \right\}\,, \tag{10.24}$$

of all images and preimages of $\mathcal{M}$. The non–wandering set Ω is the union of all points whose forward and backward trajectories remain trapped for all time.

The 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 labelled 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 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 *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.

Inspecting the fig. 10.9d 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 for the entire class of topologically similar flows.

10.6.2 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 fig. 10.10. 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 labelling 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 phase space which contains all points that share common n future and m past symbols. This region maps in a nontrivial way in the phase space, but in the symbol square its dynamics is exceedingly simple; all of its points are mapped by the decimal point shift (10.9)

10.7 on p. 234

$$\sigma(\cdots s_{-2}s_{-1}s_0.s_1s_2s_3\cdots) = \cdots s_{-2}s_{-1}s_0s_1.s_2s_3\cdots\,, \tag{10.25}$$

For example, the square [01.01] gets mapped into the rectangle $\sigma[01.01] = [010.1]$.

10.8 on p. 234

As the horseshoe mapping is a simple repetitive operation, we expect a simple relation between the symbolic dynamics labelling of the horseshoe strips, and

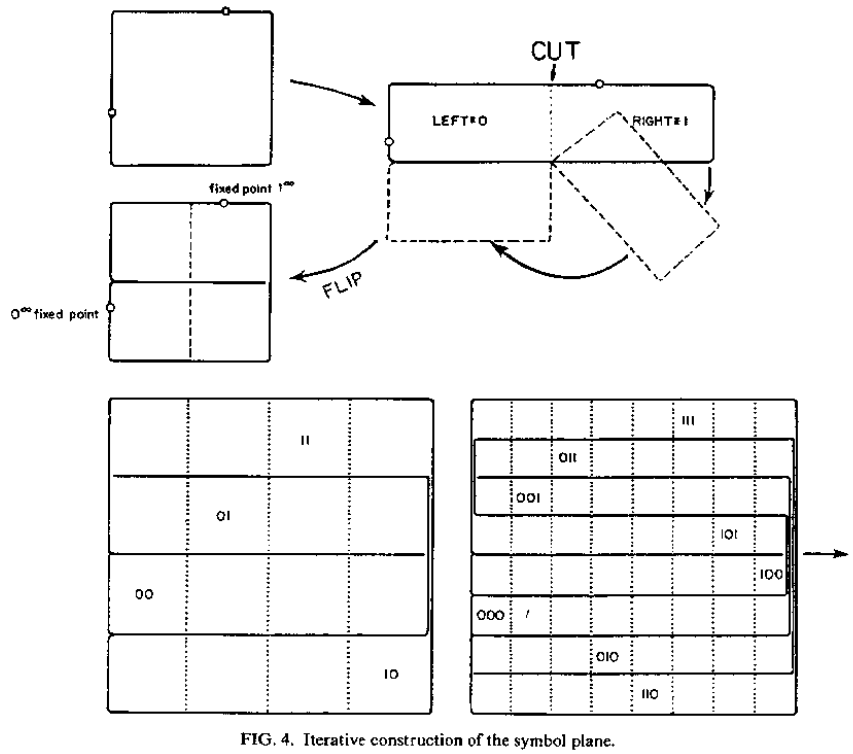


FIG. 4. Iterative construction of the symbol plane.

Figure 10.10: Kneading Danish pastry: symbol square representation of an orientation reversing once-folding map obtained by fattening the Smale horseshoe intersections of fig. 10.9 into a unit square. In the symbol square the dynamics maps rectangles into rectangles by a decimal point shift.

10.9 on p. 235 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 (10.18). This follows by inspection from fig. 10.10. In order to understand this relation between the topology of horseshoes and their symbolic dynamics, it might be helpful to backtrace to sect. 10.5.1 and work through and understand first the symbolic dynamics of 1-dimensional unimodal mappings.

10.10 on p. 236 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:

$$\begin{aligned} w_{n-1} &= \begin{cases} w_n & \text{if } s_n = 0 \\ 1 - w_n & \text{if } s_n = 1 \end{cases}, \qquad w_0 = s_0 \\ \delta(S^-) &= 0.w_0 w_{-1} w_{-2} \ldots = \sum_{n=1}^{\infty} w_{1-n}/2^n \,. \end{aligned} \tag{10.26}$$

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 (10.18).

The coordinate pair (δ, γ) maps a point (x, y) in the phase space Cantor set of fig. 10.9 into a point in the symbol square of fig. 10.10, preserving the topological ordering; (δ, γ) 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.

10.7 Pruning

> The complexity of this figure will be striking, and I shall not even try to draw it.
> H. Poincaré, describing in *Les méthodes nouvelles de la méchanique cleste* his discovery of homoclinic tangles.

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, that is, 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

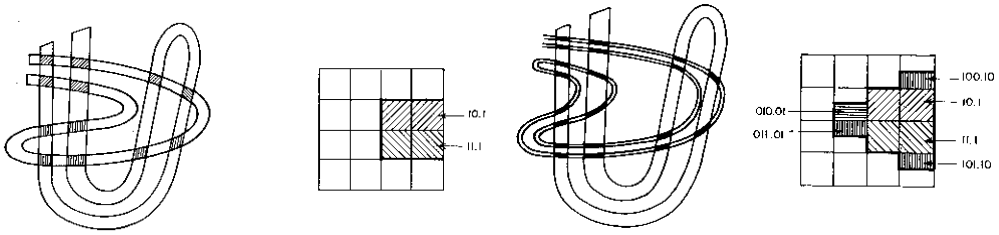


Figure 10.11: (a) An incomplete Smale horseshoe: the inner forward fold does not intersect the two rightmost backward folds. (b) The primary pruned region in the symbol square and the corresponding forbidden binary blocks. (c) An incomplete Smale horseshoe which illustrates (d) the monotonicity of the 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.

map so that the intersection of the backward and forward folds is still transverse, but no longer complete, as in fig. 10.11a. 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 fig. 10.11a 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 fig. 10.11d. 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 fig. 10.11 is inadmissible, that is, 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 (10.13). 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. 18.2 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.

10.8 Topological dynamics

So far we have established and related the *temporally* and *spatially* ordered topological dynamics for a class of "stretch & fold" dynamical systems, and given several examples of pruning of inadmissible trajectories. Now we use these results to generate the totality of admissible itineraries. This task will be relatively easy for repellers with complete Smale horseshoes and for subshifts of finite type.

10.8.1 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 labelled repeller with complete binary symbolic dynamics, we might chose to partition the phase 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 figs. 10.3 and 10.17, as well as fig. 1.7. Topologically f acts as a left shift (10.25), and its action on the rectangle [.01] is to move the decimal point to the right, to [0.1], forget the past, [.1], and land in either of the two rectangles $\{[.10], [.11]\}$. 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

10.12 on p. 237

$$\phi' = T\phi \;=\; \begin{pmatrix} T_{00,00} & 0 & T_{00,10} & 0 \\ T_{01,00} & 0 & T_{01,10} & 0 \\ 0 & T_{10,01} & 0 & T_{10,11} \\ 0 & T_{11,01} & 0 & T_{11,11} \end{pmatrix} \begin{pmatrix} \phi_{00} \\ \phi_{01} \\ \phi_{10} \\ \phi_{11} \end{pmatrix} . \tag{10.27}$$

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.

11.1 on p. 260

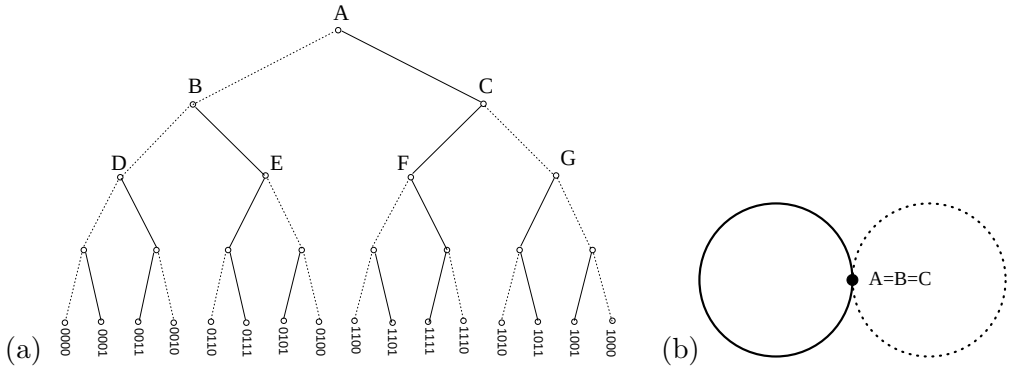
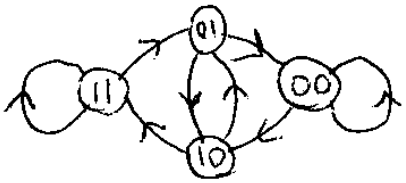


Figure 10.12: (a) The self-similarity of the complete binary symbolic dynamics represented by a binary tree (b) identification of nodes $B = A$, $C = A$ leads to the finite 1-node, 2-links Markov graph. All admissible itineraries are generated as walks on this finite Markov graph.

Figure 10.13: (a) The 2-step memory Markov graph, links version obtained by identifying nodes $A = D = E = F = G$ in fig. 10.12(a). Links of this graph correspond to the matrix entries in the transition matrix (10.27). (b) the 2-step memory Markov graph, node version.



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.

11.4 on p. 261

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 iteneraries for the case of complete binary symbolic dynamics.

11.1 on p. 260

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 two new dots, and then for the four dots, and so on. The result is the binary tree of fig. 10.12(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 fig. 10.12(a) correspond to the 16 dsitinct binary strings of length 4, and so on. By habit we have drawn the tree as the alternating binary tree of fig. 10.7, 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

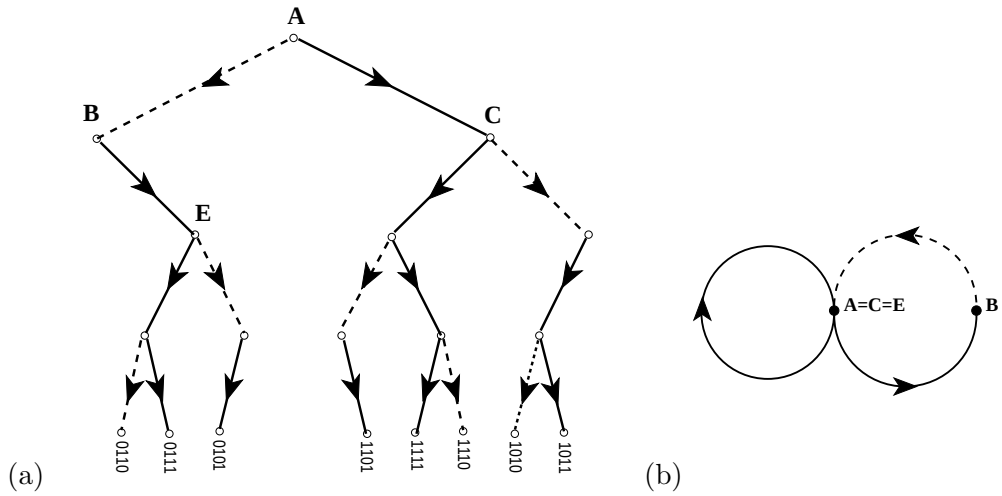


Figure 10.14: (a) The self-similarity of the _00_ pruned binary tree: trees originating from nodes C and E are the same as the entire tree. (b) Identification of nodes $A = C = E$ leads to the finite 2-node, 3-links Markov graph; as 0 is always followed by 1, the walks on this graph generate only the admissible itineraries.

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 fig. 10.12(b): any itinerary generated by the binary tree fig. 10.12(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 fig. 10.13a. The corresponding transition matrix is given by (10.27).

fast track:
chapter 11, p. 239

10.8.2 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_.

11.8 on p. 262

This situation arises, for example, in studies of the circle maps, where this kind of symbolic dynamics describes "golden mean" rotations (we shall return to this example in chapter 19). Now the admissible itineraries are enumerated

11.10 on p. 263

by the pruned binary tree of fig. 10.14(a), or the corresponding Markov graph fig. 10.14b. We recognize this as the Markov graph example of fig. 10.2.

So we can already see the main ingradients 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 (10.12) 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 begining 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.

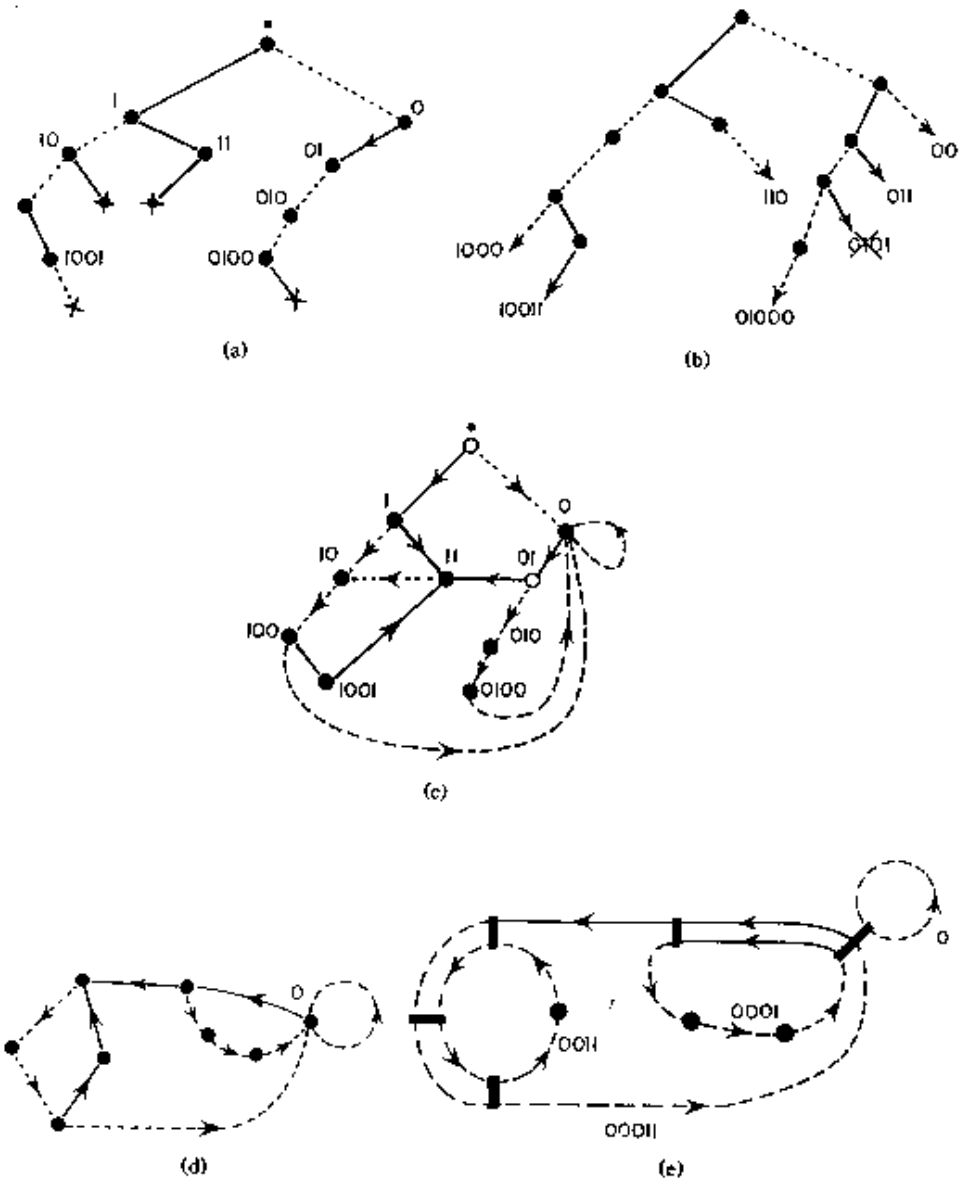


Figure 10.15: Conversion of the pruning front of fig. 10.11d into a finite Markov graph. (a) Starting with the start node ".", delineate all pruning blocks on the binary tree. A solid line stands for "1" and a dashed line for "0". Ends of forbidden strings are marked with $\times$. Label all internal 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 (11.16).

6. Clean up: check whether every node can be reached from every other node. Remove the transisent nodes, that is 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 fig. 10.11d. The pruning blocks are

10.15 on p. 238

$$[100.10], [10.1], [010.01], [011.01], [11.1], [101.10]. \tag{10.28}$$

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 (fig. 10.15a). These nodes are the potentially dangerous nodes - beginnings of blocks that might end up pruned. Add the side branches to those nodes (fig. 10.15b). 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, fig. 10.15d. 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 fig. 10.15d; the admissible bi-infinite symbol sequences are generated as all possible walks along this graph.

Commentary

Remark **10.1** Symbolic dynamics, history and good taste. For a brief history of symbolic dynamics, from J. Hadamard in 1898 onwards, see Notes to chapter 1 of Kitchens monograph [1], a very clear and enjoyable mathematical introduction to topics discussed in this chapter and the next. The

binary labeling of the once-folding map periodic points was introduced by Myrberg [13] for 1-dimensional maps, and its utility to 1-dimensional maps has been emphasized in refs. [4, **?**]. For 1-dimensional maps it is now customary to use the *R-L* notation of Metropolis, Stein and Stein [14, 18], indicating that the point x_n lies either to the left or to the right of the critical point in fig. 10.6. The symbolic dynamics of such mappings has been extensively studied by means of the Smale horseshoes, see for example ref. [7]. 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 (δ, γ) introduced in sect. 10.6.2.

Remark **10.2** Kneading theory. The admissible itineraries are studied in refs. [15, 14, 7, 6], as well as many others. We follow here the Milnor-Thurston exposition [16]. 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 (11.4), we shall not discuss it here. Baladi and Ruelle have reworked this theory in a series of papers [19, 20, 21] and in ref. [22] replaced it by a power series manipulation. The kneading theory is covered here in P. Dahlqvist's appendix E.1.

Remark **10.3** 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 [12], 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 215, rereading chapter 5 might help; for example, the Liouville operators form a Lie group (of symplectic, or canonical transformations) acting on the manifold (p, q).

Remark **10.4** Pruning fronts. The notion of a pruning front was introduced in ref. [23], and developed by K.T. Hansen for a number of dynamical systems in his Ph.D. thesis [3] and a series of papers [29]-[33]. Detailed studies of pruning fronts are carried out in refs. [24, 25, **?**]; ref. [16] is the most detailed study carried out so far. The rigorous theory of pruning fronts has been developed by Y. Ishii [26, 27] for the Lozi map, and A. de Carvalho [28] in a very general setting.

Remark **10.5** Inflating Markov graphs. In the above examples the symbolic dynamics has been encoded by labelling links in the Markov graph. Alternatively one can encode the dynamics by labelling the nodes, as in fig. 10.13, 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 (10.27).

Remark **10.6** Formal languages. Finite Markov graphs or finite automata are discussed in the present context in refs. [8, 9, 10, **?**]. They belong to the category of regular languages. A good hands-on introduction to symbolic dynamics is given in ref. [2].

Remark **10.7** The unbearable growth of Markov graphs. A construction of finite Markov partitions is described in refs. [**?**, **?**], 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. 17.3.1. 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 - fig. 10.11 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 fig. 10.15 does not generate only the "fundamental" cycles (see chapter 13), but shadowed cycles as well, such as t_{00011} in (11.16). For methods of reduction to a minimal graph, consult refs. [**?**, **?**, **?**]. 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. 11.6. 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. [24]. Still, this seems to be the best technology available, unless the reader alerts us to something superior.

Résumé

Given a partition $\mathcal{A}$ of the phase space $\mathcal{M}$, a dynamical system $(\mathcal{M}, f)$ induces topological dynamics (Σ, σ) on the space Σ of all admissible bi–infinite itineraries. The itinerary describes the time evolution of an orbit, while the symbol square describes the spatial ordering of points along the orbit. The symbol square is essential in transforming topological pruning into pruning rules for inadmissible sequences; those are implemented by constructing transition matrices and/or Markov graphs. As we shall see in the next chapter, these matrices are the simplest examples of "operators" prerequisite to developing a theory of averaging over chaotic flows.

Symbolic dynamics is the coarsest example of coarse graining, the way irreversibility enters chaotic dynamics. The exact trajectory is deterministic, and given an initial point we know (in principle) both its past and its future - its memory is infinite. In contrast, the partitioned phase space is described by the quientessentially probabilistic tools, such as the finite memory Markov graphs.

Importance of symbolic dynamics is sometime grossly unappreciated; the crucial ingredient for nice analyticity properties of zeta functions is existence of finite grammar (coupled with uniform hyperbolicity).

References

[10.1] B.P. Kitchens, Symbolic dynamics: one-sided, two-sided, and countable state Markov shifts (Springer, Berlin 1998).

[10.2] D.A. Lind and B. Marcus, *An introduction to symbolic dynamics and coding* (Cambridge Univ. Press, Cambridge 1995).

[10.3] 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).

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

[10.5] R.L. Devaney, *A First Course in Chaotic Dynamical Systems* (Addison-Wesley, Reading MA, 1992).

[10.6] R.L. Devaney, *An Introduction to Chaotic Dynamical Systems* (Addison-Wesley, Reading MA, 1987).

[10.7] J. Guckenheimer and P. Holmes, *Non-linear Oscillations, Dynamical Systems and Bifurcations of Vector Fields* (Springer, New York, 1986).

[10.8] A. Salomaa, *Formal Languages* (Academic Press, San Diego, 1973).

[10.9] J.E. Hopcroft and J.D. Ullman, *Introduction to Automata Theory, Languages, and Computation* (Addison-Wesley, Reading MA, 1979).

[10.10] D.M. Cvetković, M. Doob and H. Sachs, *Spectra of Graphs* (Academic Press, New York, 1980).

[10.11] T. Bedford, M.S. Keane and C. Series, eds., *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces* (Oxford University Press, Oxford, 1991).

[10.12] M.S. Keane, *Ergodic theory and subshifts of finite type*, in ref. [11].

[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.* **A15**, 25 (1973).

[10.15] A.N. Sarkovskii, *Ukrainian Math. J.* **16**, 61 (1964).

[10.16] 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).

[10.17] W. Thurston, *"On the geometry and dynamics of diffeomorphisms of surfaces"*, *Bull. Amer. Math. Soc. (N.S.)* **19**, 417 (1988).

[10.18] P. Collet and J.P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems* (Birkhauser, Boston, 1980).

[10.19] 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).

[10.20] V. Baladi, *"Infinite kneading matrices and weighted zeta functions of interval maps"*, *J. Functional Analysis* **128**, 226 (1995).

[10.21] D. Ruelle, *"Sharp determinants for smooth interval maps"*, Proceedings of Montevideo Conference 1995, IHES preprint (March 1995).

[10.22] V. Baladi and D. Ruelle, *"Sharp determinants"*, *Invent. Math.* **123**, 553 (1996).

[10.23] P. Cvitanović, G.H. Gunaratne and I. Procaccia, *Phys. Rev.* **A 38**, 1503 (1988).

[10.24] G. D'Alessandro, P. Grassberger, S. Isola and A. Politi, "On the topology of the Hénon Map", *J. Phys.* **A 23**, 5285 (1990).

[10.25] G. D'Alessandro, S. Isola and A. Politi, "Geometric properties of the pruning front", *Prog. Theor. Phys.* **86**, 1149 (1991).

[10.26] Y. Ishii, "Towards the kneading theory for Lozi attractors. I. Critical sets and pruning fronts", Kyoto Univ. Math. Dept. preprint (Feb. 1994).

[10.27] Y. Ishii, "Towards a kneading theory for Lozi mappings. II. A solution of the pruning front conjecture and the first tangency problem", *Nonlinearity* (1997), to appear.

[10.28] A. de Carvalho, Ph.D. thesis, CUNY New York 1995; "Pruning fronts and the formation of horseshoes", preprint (1997).

[10.29] K.T. Hansen, *CHAOS* **2**, 71 (1992).

[10.30] K.T. Hansen, *Nonlinearity* **5**

[10.31] K.T. Hansen, *Nonlinearity* **5**

[10.32] K.T. Hansen, *Symbolic dynamics III, The stadium billiard*, to be submitted to *Nonlinearity*

[10.33] K.T. Hansen, *Symbolic dynamics IV; a unique partition of maps of Hénon type*, in preparation.

Exercises

10.1 Binary symbolic dynamics. Verify that the shortest prime binary cycles of the unimodal repeller of fig. 10.6 are $\overline{0}$, $\overline{1}$, $\overline{01}$, $\overline{001}$, $\overline{011}$, $\cdots$. Compare with table 10.1. Try to sketch them in the graph of the unimodal function $f(x)$; compare ordering of the periodic points with fig. 10.7. 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 3-disk fundamental domain symbolic dynamics. Try to sketch $\overline{0}$, $\overline{1}$, $\overline{01}$, $\overline{001}$, $\overline{011}$, $\cdots$. in the fundamental domain, fig. 10.4, and interpret the symbols $\{0,1\}$ by relating them to topologically distinct types of collisions. Compare with table 10.2. Then try to sketch the location of periodic points in the Poincaré section of the billiard flow. The point of this exercise is that while in the configuration space longer cycles look like a hopeless jumble, 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 phase space coordinates which does not respect the topological organization of the flow.

10.3 Generating prime cycles. Write a program that generates all binary prime cycles up to given finite length.

10.4 Reduction of 3-disk symbolic dynamics to binary.

(a) Verify that the 3-disk cycles
$\{\overline{1\,2}, \overline{1\,3}, \overline{2\,3}\}$, $\{\overline{1\,2\,3}, \overline{1\,3\,2}\}$, $\{\overline{12\,13}$ + 2 perms.$\}$,
$\{\overline{121\,232\,313}$ + 5 perms.$\}$, $\{\overline{121\,323}$+ 2 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 10.2 by drawing them both in the full 3-disk system and in the fundamental domain, as in fig. 10.4.

(c) Optional: Can you see how the group elements listed in table 10.2 relate irreducible segments to the fundamental domain periodic orbits?

10.5 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 (10.18). This follows by inspection from the binary tree of fig. 10.7.

10.6 A Smale horseshoe. The Hénon map

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 - ax^2 + y \\ bx \end{bmatrix} \tag{10.29}$$

maps the (x, y) plane into itself - it was constructed by Hénon [1] in order to mimic the Poincaré section of once-folding map induced by a flow like the one sketched in fig. 10.5. For definitivness fix the parameters to $a = 6$, $b = -1$.

a) Draw a rectangle in the (x, y) plane such that its nth iterate by the Hénon map intersects the rectangle 2^n times.

b) Construct the inverse of the (10.29).

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 10.1. We shall compute the exact cycle points in exercise 12.13.

10.7 Kneading Danish pastry. Write down the $(x, y) \to (x, y)$ mapping that implements the baker's map of fig. 10.10, 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 (10.15)).

10.8 Kneading Danish without flipping. The baker's map of fig. 10.10 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 foldings of the symbol square.

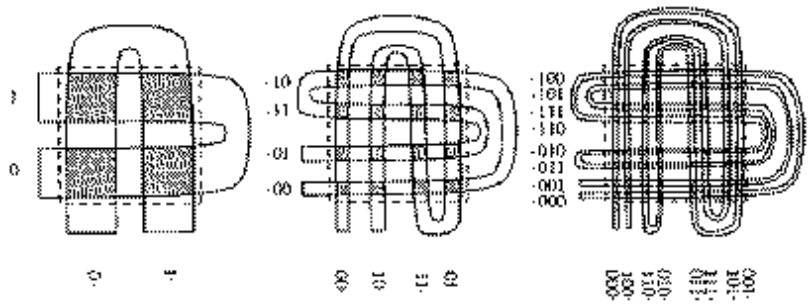


Figure 10.16: A complete Smale horseshoe iterated forwards and backwards, orientation preserving case: function f maps the dashed border square $\mathcal{M}$ into the vertical horseshoe, while the inverse map f^{-1} maps it into the horizontal horseshoe. a) One iteration, b) two iterations, c) three iterations. The non–wandering set is contained within the intersection of the forward and backward iterates (crosshatched). (from K.T. Hansen [3])

10.9 Fix this manuscript. Check whether the layers of the baker's map of fig. 10.10 are indeed ordered as the branches of the alternating binary tree of fig. 10.7. (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 fig. 10.16

(b) orientation preserving: stretch, fold downward, as in fig. 10.11

(c) orientation reversing: stretch, fold upward, flip, as in fig. 10.17

(d) orientation reversing: stretch, fold downward, flip, as in fig. 10.10,

with the corresponding four distinct binary-labelled symbol squares. For n-fold "stretch & fold" flows the labelling would be $n$ary. The intersection $\mathcal{M}_0$ for the orientation preserving Smale horseshoe, fig. 10.16a, is oriented the same way as $\mathcal{M}$, while $\mathcal{M}_1$ is oriented opposite to $\mathcal{M}$. Brief contemplation of fig. 10.10 indicates that the forward iteration strips are ordered relative to each other as the branches of the alternating binary tree in fig. 10.7.

Check the labelling for all four cases.

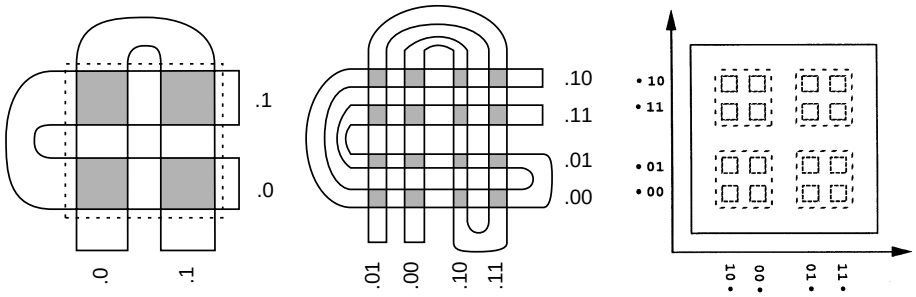
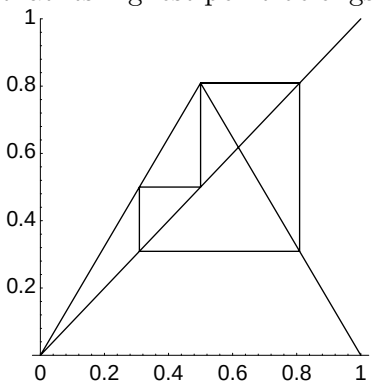


Figure 10.17: An orientation reversing Smale horseshoe map. Function $f = \{\text{stretch,fold,flip}\}$ maps the dashed border square $\mathcal{M}$ into the vertical horseshoe, while the inverse map f^{-1} maps it into the horizontal horseshoe. a) one iteration, b) two iterations, c) the non–wandering set cover by 16 rectangles, each labelled by the 2 past and the 2 future steps. (from K.T. Hansen [3])

10.10 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 fig. 10.17. $\tilde{Q}_0$ and $\tilde{Q}_1$ are oriented as Q_0 and Q_1, so the definition of the future topological coordinate γ is identical to the γ for the orientation preserving horseshoe. The inverse intersections $\tilde{Q}_0^{-1}$ and $\tilde{Q}_1^{-1}$ are oriented so that $\tilde{Q}_0^{-1}$ is opposite to Q, while $\tilde{Q}_1^{-1}$ has the same orientation as Q. Check that the past topological coordinate δ is given by

$$\begin{aligned} w_{n-1} &= \begin{cases} 1 - w_n & \text{if } s_n = 0 \\ w_n & \text{if } s_n = 1 \end{cases}, \qquad w_0 = s_0 \\ \delta(x) &= 0.w_0 w_{-1} w_{-2} \ldots = \sum_{n=1}^{\infty} w_{1-n}/2^n \,. \end{aligned} \tag{10.30}$$

10.11 "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 Λ 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 11.7 and exercise 11.9. See also exercise 11.8 and exercise 11.10.

10.12 Binary 3-step transition matrix. Construct [8×8] binary 3-step transition matrix analogous to the 2-step transition matrix (10.27). 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 $m \to \infty$.

10.13 Infinite symbolic dynamics. Let σ 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 ϵ_i are either 0 or 1. We will now define an operator $\mathcal{T}$ that acts on observables on the space of binary strings. A function a is an observable if it has bounded variation, that is, if

$$\|a\| = \sup_{\{\epsilon_i\}} |a(\epsilon_1, \epsilon_2, \ldots)| < \infty \,.$$

For these functions

$$\mathcal{T} a(\epsilon_1, \epsilon_2, \ldots) = a(0, \epsilon_1, \epsilon_2, \ldots)\sigma(0, \epsilon_1, \epsilon_2, \ldots) + a(1, \epsilon_1, \epsilon_2, \ldots)\sigma(1, \epsilon_1, \epsilon_2, \ldots) \,.$$

(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. (Hint: check if $\mathcal{T}$ is compact "trace class" is defined in appendix J.)

10.14 Time reversability.** Hamiltonian flows are time reversible. Does that mean that their Markov graphs are symmetric in all node $\to$ node links, their transition matrices are adjacency matrices, symmetric and diagonalizable, and that they have only real eigenvalues?

10.15 Heavy pruning. Implement the prunning grammar (10.28), with the pruned blocks

$$10010, 101, 01001, 01101, 111, 10110,$$

by a method of your own devising, or following the the last example of sect. 10.8 illustrated in fig. 10.15. For continuation, see exercise 11.11.

Chapter 11

Counting

> That which is crooked cannot be made straight: and that which is wanting cannot be numbered.
>
> Ecclestiastes 1.15

We are now in position to develop our first prototype application of the periodic orbit theory: cycle counting. This is the simplest illustration of raison d'etre of the 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 turn the topological dynamics of the preceding chapter into a multiplicative operation by means of transition matrices/Markov graphs, and show that the powers of a transition matrix count the distinct itineraries. 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 given by the leading zero of the characteristic determinant of the transition matrix, which is in this context called the topological zeta function. For a class of 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 the rather abstract introduction to spectral determinants commenced in chapter 8.

11.1 Counting itineraries

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 there is pruning, this is only an upper bound and explicit formulas might be hard to come by, but we still might be able to establish a lower exponential bound of form $K_n \geq Ce^{n\hat{h}}$.

Hence it is natural to characterize the growth of the number of trajectories as a function of the itinerary length by the *topological entropy*:

$$h = \lim_{n\to\infty} \frac{1}{n} \ln K_n \ . \tag{11.1}$$

We shall now relate this quantity to the eigenspectrum 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

11.1 on p. 260

$$(T^n)_{ij} = \sum_{k_1,k_2,\ldots,k_{n-1}} T_{ik_1} T_{k_1k_2} \ldots T_{k_{n-1}j}$$

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. The total number of admissible itineraries of n symbols is

$$K_n = \sum_{ij} (T^n)_{ij} = (\,1,1,\ldots,1\,)\, T^n \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} . \tag{11.2}$$

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. 11.5.2 and 11.5.3. Recommended reading if you ever have to compute lots of cycles.

T is a matrix with 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, that is, 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, and the notion is obviously inherited by the shift once we introduce a symbolic dynamics. If that is not the case, phase 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 the transition matrices of the Perron-Frobenius type.

A finite matrix T has eigenvalues $T\varphi_\alpha = \lambda_\alpha\varphi_\alpha$ and (right) eigenvectors $\{\varphi_0, \varphi_1, \cdots, \varphi_{N-1}\}$. Expressing the initial vector in (11.2) in this basis

$$T^n \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = T^n \sum_{\alpha=0}^{N-1} b_\alpha \varphi_\alpha = \sum_{\alpha=0}^{N-1} b_\alpha \lambda_\alpha^n \varphi_\alpha \,,$$

and contracting with $(1, 1, \ldots, 1)$ we obtain

$$K_n = \sum_{\alpha=0}^{N-1} c_\alpha \lambda_\alpha^n \,.$$

11.2 on p. 260

The constants c_α depend on the choice of initial and final partitions: In this example we are sandwiching T^n between the vector $(1, 1, \ldots, 1)$ and its transpose, but any other pair of vectors would do, as long as they are not orthogonal to the leading eigenvector φ_0. 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 (11.1) is given by

$$\begin{aligned} h &= \lim_{n\to\infty} \frac{1}{n} \ln c_0 \lambda_0^n \left[1 + \frac{c_1}{c_0}\left(\frac{\lambda_1}{\lambda_0}\right)^n + \cdots\right] \\ &= \ln\lambda_0 + \lim_{n\to\infty}\left[\frac{\ln c_0}{n} + \frac{1}{n}\frac{c_1}{c_0}\left(\frac{\lambda_1}{\lambda_0}\right)^n + \cdots\right] \\ &= \ln\lambda_0 \,. \end{aligned} \tag{11.3}$$

What have we learned? The transition matrix T is a one-step local 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 (11.1) 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 one dimensional lattice models with finite range interaction: the analogies with statistical mechanics will be further commented upon in chapter 15.

11.2 Topological trace formula

There are two standard ways of getting at a spectrum - by evaluating the trace $\mathrm{tr}\,T^n = \sum \lambda_\alpha^n$, or by evaluating the determinant $\det(1 - zT)$. We start by

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

Table 11.1: 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 10.1 and 11.2. (L. Rondoni)

evaluating the trace of transition matrices.

Consider an M-step memory transition matrix, like the 1-step memory example (10.27). 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 (check (10.27) and exercise 10.12). We can even take $M \to \infty$, in which case the contributing partitions are shrunk to the fixed points, $\mathrm{tr}\, T = T_{\overline{0},\overline{0}} + T_{\overline{1},\overline{1}}$.

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 walk shifts the symbolic label by one label; the trace ensures that the walk closes into 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 $[8\times8]$ transition matrix $T_{s_1s_2s_3,s_0s_1s_2}$ version of (10.27), or any refined partition $[2^n\times2^n]$ 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 labelled non–wandering set the first few traces are given by (consult tables 10.1 and 11.1)

10.12 on p. 237

$$
\begin{aligned}
z\,\mathrm{tr}\, T &= t_0 + t_1, \\
z^2 \mathrm{tr}\, T^2 &= t_0^2 + t_1^2 + 2t_{10}, \\
z^3 \mathrm{tr}\, T^3 &= t_0^3 + t_1^3 + 3t_{100} + 3t_{101}, \\
z^4 \mathrm{tr}\, T^4 &= t_0^4 + t_1^4 + 2t_{10}^2 + 4t_{1000} + 4t_{1001} + 4t_{1011}.
\end{aligned} \tag{11.4}
$$

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 nth order trace (11.4) 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

$$N_n = \mathrm{tr}\, T^n = \sum_{n_p|n} n_p t_p^{n/n_p} = \sum_p n_p \sum_{r=1}^{\infty} \delta_{n,n_p r} t_p^r \,. \tag{11.5}$$

Here $m|n$ means that m is a divisor of n, and we have taken $z = 1$ so $t_p = 1$ 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 sum, we introduce the generating function for numbers of periodic points

$$\sum_{n=1}^{\infty} z^n N_n = \mathrm{tr}\, \frac{zT}{1-zT} \,. \tag{11.6}$$

Substituting (11.5) 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 still another example of a trace formula, the *topological trace formula*

$$\sum_{\alpha=0} \frac{z\lambda_\alpha}{1-z\lambda_\alpha} = \sum_p \frac{n_p t_p}{1-t_p} \,. \tag{11.7}$$

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 (11.6) is a discrete version of the Laplace transform, see chapter 7, and the resolvent on the left hand side is the antecedent of the more sophisticated trace formulas (7.9), (7.19) and (22.3). We shall now use this result to compute the spectral determinant of the transition matrix.

11.3 Determinant of a graph

Our next task is to determine the zeros of the *spectral determinant* of an $[MxM]$ transition matrix

10.14 on p. 238

$$\det(1 - zT) = \prod_{\alpha=0}^{M-1} (1 - z\lambda_\alpha) \,. \tag{11.8}$$

We could now proceed to diagonalize T on a computer, and get this over with. Nevertheless, it pays 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 13 for arbitrary dynamical averages. For T a finite matrix (11.8) 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 ∞, and for that reason we prefer to refer to (11.8) 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 of chapter 1:

1.3 on p. 32

$$\begin{aligned}\det(1-zT) &= \exp(\mathrm{tr}\,\ln(1-zT)) = \exp\left(-\sum_{n=1}\frac{z^n}{n}\mathrm{tr}\,T^n\right)\\ &= 1-z\,\mathrm{tr}\,T-\frac{z^2}{2}\left((\mathrm{tr}\;T)^2-\mathrm{tr}\,(T^2)\right)-\ldots \end{aligned} \tag{11.9}$$

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. Coefficients of z^n, $n > M$ vanish *exactly*.

We now proceed to relate the determinant in (11.9) to the corresponding Markov graph of chapter **??**: to this end we start by the usual algebra textbook expression

$$\det(1-zT) = \sum_{\{\pi\}}(-1)^{P_\pi}(1-zT)_{1,\pi_1}\cdot(1-zT)_{2,\pi_2}\cdots(1-zT)_{M,\pi_M} \tag{11.10}$$

where once again we suppose T is an $[M\times M]$ finite matrix, $\{\pi\}$ denotes the set of permutations of M symbols, π_k is what k is permuted into by the permutation k, and P_π is the parity of the considered permutation. The right hand side of (11.10) 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

$$(-z)^n(-1)^{P_{\tilde\pi}}T_{\eta_1,\tilde\pi_{\eta_1}}\cdots T_{\eta_n,\tilde\pi_{\eta_n}} \tag{11.11}$$

where $\tilde\pi$ is the permutation of the subset of n distinct symbols $\eta_1\ldots\eta_n$ indexing T matrix elements. As in (11.4), we refer to any combination $t_i = T_{\eta_1\eta_2}T_{\eta_2\eta_3}\cdots T_{\eta_k\eta_1}$, $c=\eta_1,\eta_2,\cdots,\eta_k$ fixed, as a *local trace* associated with a closed loop c on the Markov graph. Each term of form (11.11) 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 η_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 (11.11) 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 2 (we must thus have $n = 2j + i$). The parity of the permutation gives $(-1)^j$ and the first factor in (11.11) 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:

11.3 on p. 260

> The characteristic polynomial of a transition matrix/Markov graph is given by the sum of all possible partitions π of the graph into products of non-intersecting loops, with each loop trace t_p carrying a minus sign:

$$\det(1 - zT) = \sum_{k=0}^{f} {\sum_{\pi}}' (-1)^k t_{p_1} \cdots t_{p_k} \tag{11.12}$$

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 (11.12) 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 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. 11.4, p. 247

11.3.1 Topological polynomials: learning by examples

The above definition of the determinant in terms of traces is most easily grasped by a working through a few examples. The complete binary dynamics Markov

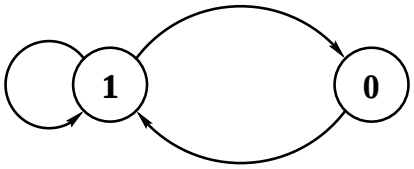


Figure 11.1: The golden mean pruning rule Markov graph, see also fig. 10.14

graph of fig. 10.12(b) is a little bit too simple, but anyway, let us start humbly; there are only two non-intersecting loops, yielding

$$\det(1 - zT) = 1 - t_0 - t_1 = 1 - 2z\,. \tag{11.13}$$

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. Similarly, for complete symbolic dynamics of N symbols the Markov graph has one node and N links, yielding

$$\det(1 - zT) = 1 - Nz\,, \tag{11.14}$$

whence the topological entropy $h = \ln N$.

A more interesting example is the "golden mean" pruning of fig. 11.1. There is only one grammar rule, that a repeat of symbol o is forbidden. The non-intersecting loops are of length 1 and 2, so the topological polynomial is given by

11.4 on p. 261

$$\det(1 - zT) = 1 - t_1 - t_{01} = 1 - z - z^2\,. \tag{11.15}$$

The leading root of this polynomial is the golden mean, so the entropy (11.3) is the logarithm of the golden mean, $h = \ln \frac{1+\sqrt{5}}{2}$.

Finally, the non-self-intersecting loops of the Markov graph of fig. 10.15(d) are indicated in fig. 10.15(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:

11.11 on p. 263

$$\det(1 - T) = 1 - t_0 - t_{0011} - t_{0001} - t_{00011} + t_0 t_{0011} + t_{0011} t_{0001} \tag{11.16}$$

With $t_p = z^{n_p}$, where n_p is the length of the p-cycle, the smallest root of

11.12 on p. 263

$$0 = 1 - z - 2z^4 + z^8 \tag{11.17}$$

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 $h = \ln 2 = 0.693\ldots$

in depth:
sect. L.1, p. 725

11.4 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 the idea is central to the whole treatise: the determinant is then defined by its *cumulant* expansion (11.9)

1.3
on p. 32

$$\det(1 - zT) = 1 - \sum_{n=1}^{\infty} \hat{c}_n z^n \,. \tag{11.18}$$

For finite dimensional matrices the expansion is a finite polynomial, and (11.18) 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 (11.8) in terms of cycles, substitute (11.5) into (11.18) and sum over the repeats of prime cycles

$$\det(1 - zT) = \exp\left(-\sum_{p}\sum_{r=1}^{\infty}\frac{t_p^r}{r}\right) = \prod_{p}(1 - t_p) \,. \tag{11.19}$$

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

$$1/\zeta_{\rm top} = \prod_{p}(1 - z^{n_p}) = 1 - \sum_{n=1} \hat{c}_n z^n \,. \tag{11.20}$$

Counting cycles amounts to giving each admissible prime cycle p weight $t_p = z^{n_p}$ and expanding the Euler product (11.20) 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 deriving it until chapter 13.

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 (11.12) is finite, and we refer to this expansion as the *topological polynomial* of order $\leq N$. The power of defining a determinant by the cumulant expansion is that it works even when the partition is infinite, $N \to \infty$; an example is given in sect. 11.6, and many more later on.

fast track:
sect. 11.6, p. 252

11.4.1 Topological zeta function for flows

We now apply the method we used in deriving (7.19) to the problem of deriving the topological zeta functions for flows. By analogy to (7.17), the time-weighted density of prime cycles of period t is

$$\Gamma(t) = \sum_p \sum_{r=1} T_p \delta(t - rT_p) \,. \tag{11.21}$$

A Laplace transform smoothes the sum over Dirac delta spikes and yields the *topological trace formula*

$$\sum_p \sum_{r=1} T_p \int_{0_+}^{\infty} dt\, e^{-st}\, \delta(t - rT_p) = \sum_p T_p \sum_{r=1}^{\infty} e^{-sT_p r} \tag{11.22}$$

and the *topological zeta function* for flows:

$$\begin{aligned} 1/\zeta_{\rm top}(s) &= \prod_p \left(1 - e^{-sT_p}\right) \\ \sum_p T_p \sum_{r=1}^{\infty} e^{-sT_p r} &= -\frac{\partial}{\partial s} \ln 1/\zeta_{\rm top}(s) \,. \end{aligned} \tag{11.23}$$

This is the continuous time version of the discrete time topological zeta function (11.20) for maps; its leading zero $s = -h$ yields the topological entropy for a flow.

11.5 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.

11.5.1 Counting periodic points

N_n, the number of periodic points of period n can be computed from (11.18) and (11.6) as a logarithmic derivative of the topological zeta function

$$\begin{aligned}\sum_{n=1} N_n z^n &= \mathrm{tr}\left(-z\frac{d}{dz}\ln(1-zT)\right) = -z\frac{d}{dz}\ln\det\,(1-zT) \\ &= \frac{-z\frac{d}{dz}1/\zeta_{\rm top}}{1/\zeta_{\rm top}}\,. \end{aligned} \tag{11.24}$$

We see that the trace formula (11.7) diverges at $z \to e^{-h}$, as the denominator has a simple zero there.

As a check of formula (11.18) 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 (11.18) we verify (11.14). The logarithmic derivative formula (11.24) in this case does not buy us much either, we recover

$$\sum_{n=1} N_n z^n = \frac{Nz}{1-Nz}\,.$$

However, consider instead the nontrivial pruning of fig. 10.15(e). Substituting (11.17) we obtain

$$\sum_{n=1} N_n z^n = \frac{z+8z^4-8z^8}{1-z-2z^4+z^8}\,. \tag{11.25}$$

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 $n > 8$.

11.5.2 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

$$N_n = \sum_{d|n} dM_d \,, \tag{11.26}$$

where N_n equals $\operatorname{tr} T^n$. The number of prime cycles can be computed recursively

$$M_n = \frac{1}{n}\left(N_n - \sum_{d|n}^{d<n} dM_d \right) \,,$$

11.13 on p. 264 or by the *Möbius inversion formula*

$$M_n = n^{-1} \sum_{d|n} \mu\left(\frac{n}{d}\right) N_d \,. \tag{11.27}$$

11.14 on p. 264 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 11.2. The number of *prime* cycles follows by Möbius inversion (11.27).

11.5.3 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 $N_n = \operatorname{tr} T^n$. The pruning of self-bounces eliminates

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

Table 11.2: Number of prime cycles for various alphabets and grammars up to length 10. The first column gives the cycle length, the second the formula (11.27) for the number of prime cycles for complete N-symbol dynamics, columns three through five give the numbers for $N = 2, 3$ and 4.

the diagonal entries, $T_{N-disk} = T_c - \mathbf{1}$, so the number of the N-disk periodic points is

$$N_n = \mathrm{tr}\, T^n_{N-disk} = (N-1)^n + (-1)^n(N-1) \tag{11.28}$$

(here T_c is the complete symbolic dynamics transition matrix (10.3)). For the N-disk pruned case (11.28) Möbius inversion (11.27) yields

$$\begin{aligned} M_n^{N-disk} &= \frac{1}{n}\sum_{d|n}\mu\left(\frac{n}{d}\right)(N-1)^d + \frac{N-1}{n}\sum_{d|n}\mu\left(\frac{n}{d}\right)(-1)^d \\ &= M_n^{(N-1)} \quad \text{for} \quad n > 2\ . \end{aligned} \tag{11.29}$$

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 11.2.

11.5.4 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 (11.29)). 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 $(1-z^2t_{12})(1-z^2t_{13})(1-z^2t_{23})$:

11.17 on p. 265

11.18 on p. 265

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 11.3: 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 17 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.

$$\begin{aligned} 1/\zeta_{3-disk} &= (1-2z)\frac{(1-z^2)^3}{(1-z)^2(1-z^2)} \\ &= (1-2z)(1+z)^2 = 1-3z^2-2z^3\,. \end{aligned} \tag{11.30}$$

The factorization reflects the underlying 3-disk symmetry; we shall rederive it in (17.25). As we shall see in chapter 17, symmetries lead to factorizations of topological polynomials and topological zeta functions.

11.19 on p. 266

The example of exercise 11.19 with the alphabet $\{a, cb^k;\ \overline{b}\}$ is more interesting. 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 $(1-t_b) = (1-z)$:

$$1/\zeta_{\rm top} = (1-z)\left(1-z-\frac{z}{1-z}\right) = 1-3z+z^2$$

11.22 on p. 267

11.6 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

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	

Table 11.4: List of the 4-disk prime cycles up to length 8. The meaning of the symbols is the same as in table 11.3. 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.

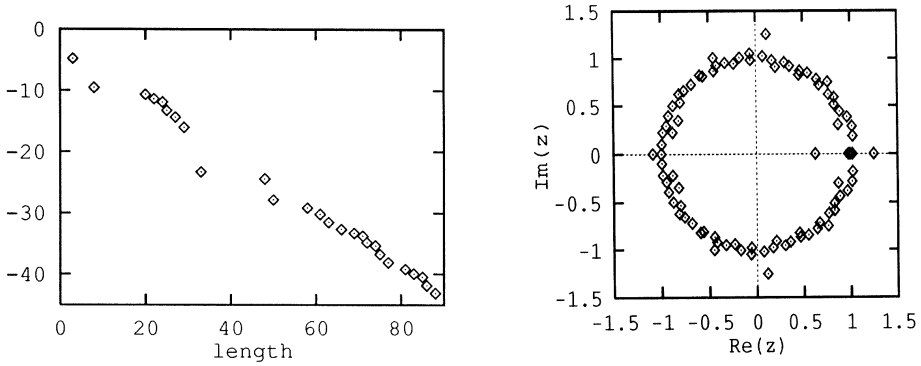


Figure 11.2: (a) The logarithm of the difference between the leading zero of the finite polynomial approximations to topological zeta function and our best estimate, as a function of the length for the quadratic map $A = 3.8$. (b) The 90 zeroes of the characteristic polynomial for the quadratic map $A = 3.8$ approximated by symbolic strings up to length 90. (from ref. [3])

- 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" (see sect. 10.5.2) of the critical point $x = 1/2$ is

$$K = 10110111101101111010111101111110\ldots$$

where the symbolic dynamics is defined by the partition of fig. 10.6. How this kneading sequence is converted into a series of pruning rules is a dark art, relegated to appendix E.1 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 (or, much easier, following the algorithm of appendix E.1) yields a Markov graph with 29 nodes and the characteristic polynomial

$$\begin{aligned} 1/\zeta^{(29)}_{\rm top} \;=\; & 1 - z^1 - z^2 + z^3 - z^4 - z^5 + z^6 - z^7 + z^8 - z^9 - z^{10} \\ & + z^{11} - z^{12} - z^{13} + z^{14} - z^{15} + z^{16} - z^{17} - z^{18} + z^{19} + z^{20} \\ & - z^{21} + z^{22} - z^{23} + z^{24} + z^{25} - z^{26} + z^{27} - z^{28}\,. \end{aligned} \tag{11.31}$$

The smallest real root of this approximate topological zeta function is

$$z = 0.62616120\ldots \tag{11.32}$$

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:

$$h = -\ln 0.62616130424685\ldots = 0.46814726655867\ldots. \tag{11.33}$$

Fig. 11.2 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 (11.33), plotted as a function of the length of the stable periodic orbit. The error of the estimate (11.32) 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 fig. 11.2(b) 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 $\lambda_1/\lambda_0 = 1/e^h = e^{-h}$.

11.7 Shadowing

The topological zeta function is a pretty function, but the infinite product (11.19) should make you pause. For finite transfer matrices the left hand side is a determinant of a finite matrix, therefore a finite polynomial; but the right hand side is 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 (13.5) of chapter 13; 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 8.2. As we have already argued in sect. 1.4.4 and appendix I.1.2, 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 (11.12), with longer cycles contributing only small "curvature" corrections.

These exact or nearly exact cancellations 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 fig. 11.2(b).

One striking aspect of the pruned cycle expansion (11.31) compared to the trace formulas such as (11.6) 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 (11.18) increases the *analyticity domain*: the pole in the trace (11.7) at $z = e^{-h}$ is promoted to a smooth zero of the spectral determinant with a larger radius of convergence.

A detailed discussion of the radius of convergence is given in appendix E.1.

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. 10.6.1), 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 18.

You might well ask: What is wrong with computing an entropy from (11.1)? Does all this theory buy us anything? If we count K_n level by level, we ignore the self-similarity of the pruned tree - examine for example fig. 10.14, or the cycle expansion of (11.25) - 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 (11.3). The determinant (11.8) 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.

Commentary

Remark **11.1** "Entropy". The ease with which the topological entropy can be motivated obscures the fact that our definition does not lead to an invariant of the dynamics, as the choice of symbolic dynamics is largely arbitrary: the same caveat applies to other entropies discussed in chapter 15, and to get proper invariants one is forced to evaluating a supremum over all

possible partitions. The key mathematical point that eliminates the need of such a variational search is the existence of *generators*, *i.e.* partitions that under dynamics are able to probe the whole phase 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 phase space generated by Ω, and consider the partition built upon all possible intersectiond of sets $\phi^k(\beta_i)$, where ϕ is dynamical evolution, β_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] and chapter 15.

Remark **11.2** Perron-Frobenius matrices. For a proof of Perron theorem on the leading eigenvalue see ref. [2]. Ref. [3], sect. A4.1 contains a clear discussion of the spectrum of the transition matrix.

Remark **11.3** 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. [4, 5, 6].

Remark **11.4** T is not trace class. Note to the erudite reader: the transition matrix T (in the infinite partition limit (11.18)) is *not* trace class in the sense of appendix J. Still the trace is well defined in the $n \to \infty$ limit.

Remark **11.5** Artin-Mazur zeta functions. Motivated by A. Weil's zeta function for the Frobenius map [7], Artin and Mazur [13] introduced the zeta function (11.20) that counts periodic points for diffeomorphisms (see also ref. [8] for their evaluation for maps of the interval). Smale [9] conjectured rationality of the zeta functions for Axiom A diffeomorphisms, later proved by Guckenheimer [10] and Manning [11]. See remark 8.4 on page 160 for more zeta function history.

Remark **11.6** Ordering periodic orbit expansions. In sect. 13.4 we will introduce an alternative way of hierarchically organising 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.

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 (11.20) of the topological zeta function (that is, the spectral determinant of the transition matrix):

$$1/\zeta_{\rm top}(z) \;=\; 1 - \sum_{k=1} \hat{c}_k z^k \,.$$

For subshifts of finite type, the transition matrix is finite, and the topological zeta function is a finite polynomial evaluated by the loop expansion (11.12) 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 (11.1), 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 were the dynamical zeta functions that derived in chapter 7, the weighted generalizations 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. 11.6, and fig. 11.2 in particular.

References

[11.1] V.I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics*, (Addison-Wesley, Redwood City 1989)

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

[11.3] J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena*, (Clarendon Press, Oxford 1996)

[11.4] A. Salomaa, *Formal Languages*, (Academic Press, San Diego 1973)

[11.5] J.E. Hopcroft and J.D. Ullman, *Introduction to Automata Theory, Languages and Computation*, (Addison-Wesley, Reading Ma 1979)

[11.6] D.M. Cvektović, M. Doob and H. Sachs, *Spectra of Graphs*, (Academic Press, New York 1980)

[11.7] A. Weil, *Bull.Am.Math.Soc.* **55**, 497 (1949)

[11.8] 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)

[11.9] S. Smale, *Ann. Math.*, **74**, 199 (1961).

[11.10] J. Guckenheimer, *Invent.Math.* **39**, 165 (1977)

[11.11] A. Manning, *Bull.London Math.Soc.* **3**, 215 (1971)

Exercises

11.1 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 11.3.

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 $\{0, 1\}$.

11.2 Sum of A_{ij} is like a trace. Let A be a matrix with eigenvalues λ_k. Show that

$$\Gamma_n = \sum_{i,j} [A^n]_{ij} = \sum_k c_k \lambda_k^n \,.$$

(a) Use this to show that $\ln |\mathrm{tr}\, A^n|$ and $\ln |\Gamma_n|$ have the same asymptotic behavior as $n \to \infty$, that is, their ratio converges to one.

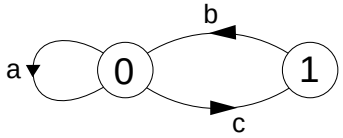
(b) Do eigenvalues λ_k need to be distinct, $\lambda_k \neq \lambda_l$ for $k \neq l$?

11.3 Loop expansions. Prove by induction the sign rule in the determinant expansion (11.12):

$$\det (1 - z\mathbf{T}) = \sum_{k \geq 0} \sum_{p_1 + \cdots + p_k} (-1)^k t_{p_1} t_{p_2} \cdots t_{p_k} \,.$$

11.4 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?

11.5 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, $\cdots$.

11.6 Dynamical zeta functions from Markov graphs. Extend sect. 11.3 to evaluation of dynamical zeta functions for piecewise linear maps with finite Markov graphs. This generalizes the results of exercise 8.2.

11.7 "Golden mean" pruned map. Continuation of exercise 10.11: 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 11.9. See also exercise 11.10.

11.8 Alphabet {0,1}, prune _00_ . The Markov diagram fig. 10.14(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 (11.12) becomes

$$\begin{aligned} 1/\zeta &= (1-t_1)(1-t_2)(1-t_{12})(1-t_{112})\dots \\ &= 1-t_1-t_2-(t_{12}-t_1t_2)-(t_{112}-t_{12}t_1)-(t_{122}-t_{12}t_2)\dots \end{aligned} \tag{11.34}$$

In the original binary alphabet this corresponds to:

$$\begin{aligned} 1/\zeta &= 1-t_1-t_{10}-(t_{110}-t_1t_{10}) \\ &\quad -(t_{1110}-t_{110}t_1)-(t_{11010}-t_{110}t_{10})\dots \end{aligned} \tag{11.35}$$

This symbolic dynamics describes, for example, circle maps with the golden mean winding number, see chapter 19. For unimodal maps this symbolic dynamics is realized by the tent map of exercise 11.7.

11.9 Spectrum of the "golden mean" pruned map. (medium - Exercise 11.7 continued)

(a) Determine an expression for $\mathrm{tr}\,\mathcal{L}^n$, the trace of powers of the Perron-Frobenius operator (5.10) for the tent map of exercise 11.7.

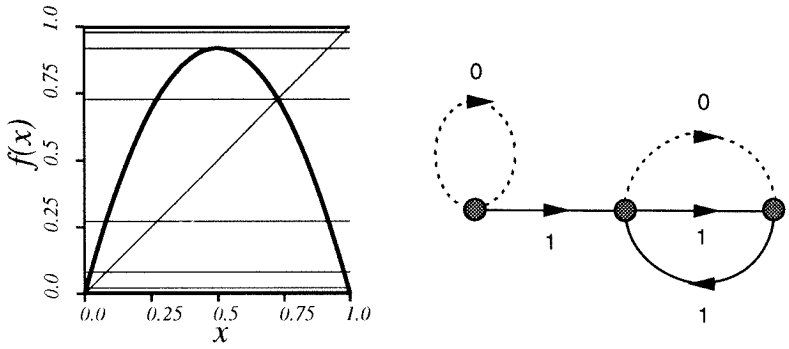


Figure 11.3: (a) A unimodal map for which the critical point maps into the right hand fixed point in three iterations, and (b) the corresponding Markov graph (Kai T. Hansen).

(b) Show that the spectral determinant for the Perron-Frobenius operator is

$$\det(1-z\mathcal{L}) = \prod_{k \text{ even}} \left(1 + \frac{z}{\Lambda^{k+1}} - \frac{z^2}{\Lambda^{2k+2}}\right) \prod_{k \text{ odd}} \left(1 + \frac{z}{\Lambda^{k+1}} + \frac{z^2}{\Lambda^{2k+2}}\right) . \quad (11.36)$$

11.10 A unimodal map example. Consider a unimodal map of fig. 11.3(a) 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 fig. 11.3(b).

(Kai T. Hansen)

11.11 Heavy pruning. (continuation of exercise 10.15.) Implement the grammar (10.28) by verifying all steps in the construction outlined in fig. 10.15. Verify the entropy estimate (11.17). Perhaps count admissible trajectories up to some length of 5-10 symbols by your own method (generate all binary sequences, throw away the bad ones?), check whether this converges to the h value claimed in the text.

11.12 Glitches in shadowing.** Note that the combination t_{00011} minus the "shadow" $t_0 t_{0011}$ in (11.16) cancels exactly, and does not contribute to the topological polynomial (11.17). Are you able to construct a smaller Markov graph than fig. 10.15(e)?

11.13 Whence Möbius function? To understand where the Möbius function comes from consider the function

$$f(n) = \sum_{d|n} g(d) \tag{11.37}$$

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*

$$g(n) = \sum_{d|n} \mu(n/d) f(d) \tag{11.38}$$

11.14 Counting prime binary cycles. In order to get comfortable with Möbius inversion reproduce the results of the second column of table 11.2.

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.

11.15 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 (10.15) 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 (11.4) collapse to $tr(T^k) = (t_0 + t_1)^k$, and $1/\zeta$ is simply

$$\prod_p (1 - t_p) = 1 - t_0 - t_1 \tag{11.39}$$

Substituting (11.39) into the identity

$$\prod_p (1 + t_p) = \prod_p \frac{1 - {t_p}^2}{1 - t_p}$$

we obtain

$$\begin{aligned} \prod_p (1 + t_p) &= \frac{1 - t_0^2 - t_1^2}{1 - t_0 - t_1} = 1 + t_0 + t_1 + \frac{2t_0 t_1}{1 - t_0 - t_1} \\ &= 1 + t_0 + t_1 + \sum_{n=2}^{\infty} \sum_{k=1}^{n-1} 2 \binom{n-2}{k-1} t_0^k t_1^{n-k}. \end{aligned} \tag{11.40}$$

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

$$\begin{aligned} M_{1,0} &= M_{1,1} = 1 \\ nM_{n,k} &= \sum_{m \left| \frac{n}{k} \right.} \mu(m) \binom{n/m}{k/m}, \quad n \geq 2\,, k = 1, \ldots, n-1 \end{aligned}$$

where the sum is over all m which divide both n and k.

11.16 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?

11.17 4-disk pinball topological polynomial. Show that the 4-disk pinball topological polynomial (the pruning affects only the fixed points and the 2-cycles) is given by

$$\begin{aligned} 1/\zeta_{4-disk} &= (1-3z)\frac{(1-z^2)^6}{(1-z)^3(1-z^2)^3} \\ &= (1-3z)(1+z)^3 = 1-6z^2-8z^3-3z^4\ . \end{aligned} \tag{11.41}$$

11.18 N-disk pinball topological polynominal. Show that for an N-disk pinball, the topological polynominal is given by

$$\begin{aligned} 1/\zeta_{N-disk} &= (1-(N-1)z)\,\frac{(1-z^2)^{N(N-1)/2}}{(1-z)^{N-1}(1-z^2)^{(N-1)(N-2)/2}} \\ &= (1-(N-1)z)\,(1+z)^{N-1}\ . \end{aligned} \tag{11.42}$$

The topological polynomial has a root $z^{-1} = N-1$, as we already know it should from (11.28) or (11.14). We shall see in sect. 17.4 that the other roots reflect the symmetry factorizations of zeta functions.

11.19 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 (11.12) becomes

$$\begin{aligned} 1/\zeta &= (1-t_a)(1-t_b)(1-t_c)(1-t_{cb})(1-t_{ac})(1-t_{cbb})\dots \\ &= 1-t_a-t_b-t_c+t_at_b-(t_{cb}-t_ct_b)-(t_{ac}-t_at_c)-(t_{cbb}-t_{cb}t_b)\dots \end{aligned}$$

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.

11.20 Alphabet {0,1}, prune n repeats. of "0" $_000\dots00_$.

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 (11.12) becomes

$$1/\zeta = 1-t_1-t_2\dots-t_n-(t_{12}-t_1t_2)\dots-(t_{1n}-t_1t_n)\dots \tag{11.43}$$

.

11.21 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, sect. 10.7.

Show that the topological zeta function is given by

$$1/\zeta = (1-t_0)(1-t_1-t_2-t_{23}-t_{113}) \tag{11.44}$$

with the unrestricted 4-letter alphabet {1, 2, $\underline{23}$, $\underline{113}$}. Here 2, 3, refer to 10, 100 respectively, as in exercise 11.20.

11.22 Alphabet {0,1}, prune _1000_, _00100_, _01100_, _10011_. This example of pruning we shall use in sect. **??**. 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 11.21) to the alphabet $\{\underline{21^k}, \underline{23}, \underline{21^k113}; \overline{1}, \overline{0}\}$, and the cycle expansion

$$1/\zeta = (1-t_0)(1-t_1-t_2-t_{23}+t_1t_{23}-t_{2113}) \tag{11.45}$$

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 polynomial is

$$1/\zeta_{\rm top} = (1-z)(1-z-z^2-z^5+z^6-z^7) \tag{11.46}$$

and check that it yields the exact value of the entropy $h = 0.522737642\ldots$.

Chapter 12

Fixed points, and how to get them

(F. Christiansen)

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 7 that cycles are the necessary ingredient for evaluation of spectra of evolution operators. In chapter **??** 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.

fast track:
chapter 5, p. 97

A *prime* cycle p of period T_p is a single traversal of the orbit, so our task will be to find a cycle point $x \in p$ and the shortest time $T = T_p$ for which (12.1) has a solution. A cycle point of a flow which crosses a Poincaré section n_p times is a fixed point of the f^{n_p} iterate of the Poincaré section return map, 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 stability point, so it will suffice to solve (12.1) at a single cycle point.

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 are considerably more robust and safer.

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 10.1. The relations between the temporal symbol sequences and the spatial layout of the topologically distinct regions of the phase space discussed in chapter **??** 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 illustrate this by considering 1-dimensional and d-dimensional maps.

12.1 One-dimensional mappings

12.1.1 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. Fig. 12.1 shows such path to the $\overline{01}$-cycle of the logistic map.

12.13 on p. 290

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 12.13. 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 $f^T(x) = x$.

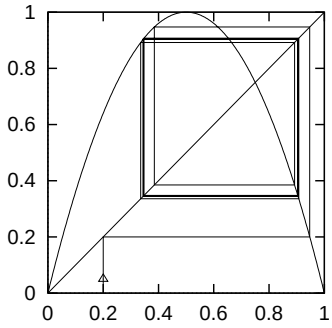


Figure 12.1: The inverse time path to the $\overline{01}$-cycle of the logistic map f(x)=4x(1-x) from an initial guess of x=0.2. At each inverse iteration we chose the 0, respectively 1 branch.

12.1.2 Newton's method

Newton's method for finding solutions of $F(x) = 0$ works as a simple linearization around a starting guess x_0:

$$F(x) \approx F(x_0) + F'(x_0)(x - x_0). \tag{12.2}$$

An approximate solution x_1 of $F(x) = 0$ is

$$x_1 = x_0 - F(x_0)/F'(x_0). \tag{12.3}$$

The approximate solution can then be used as a new starting guess in an iterative process. A fixed point of a map f is a solution to $F(x) = x - f(x) = 0$. We determine x by iterating

$$\begin{aligned} x_m &= g(x_{m-1}) = x_{m-1} - F(x_{m-1})/F'(x_{m-1}) \\ &= x_{m-1} - \frac{1}{1 - f'(x_{m-1})}(x_{m-1} - f(x_{m-1}))\,. \end{aligned} \tag{12.4}$$

Privided 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 inital guess, the Newton-Raphson iteration will converge super-exponentially fast. In fact, as is illustrated by fig. 12.2, in the typical case the number of significant digits of the accuracy of x estimate doubles with each iteration.

12.1.3 Multipoint shooting method

Periodic orbits of length n are fixed points of f^n so in principle we could use the simple Newton's 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 phase-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.

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 phase space discussed in chapter **??** 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's method now takes the form of

$$\frac{d}{dx}F(x)(x' - x) = -F(x), \tag{12.5}$$

where $\frac{d}{dx}F(x)$ is an $[n \times n]$ matrix:

$$\tfrac{d}{dx}F(x) \;=\; \begin{pmatrix} 1 & & & & -f'(x_n) \\ -f'(x_1) & 1 & & & \\ & \cdots & 1 & & \\ & & \cdots & 1 & \\ & & & -f'(x_{n-1}) & 1 \end{pmatrix} . \tag{12.6}$$

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. Let us take it step by step for a period 3 cycle. Initially the setup for the Newton step looks like this:

$$\begin{pmatrix} 1 & 0 & -f'(x_3) \\ -f'(x_1) & 1 & 0 \\ 0 & -f'(x_2) & 1 \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix} = \begin{pmatrix} -F_1 \\ -F_2 \\ -F_3 \end{pmatrix} , \tag{12.7}$$

where $\delta_i = x'_i - x_i$ is the correction of our guess for a solution and where $F_i = x_i - f(x_{i-1})$. First we eliminate the below 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. We then have

$$\begin{pmatrix} 1 & 0 & -f'(x_3) \\ 0 & 1 & -f'(x_1)f'(x_3) \\ 0 & 0 & 1 - f'(x_2)f'(x_1)f'(x_3) \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix} = \begin{pmatrix} -F_1 \\ -F_2 - f'(x_1)F_1 \\ -F_3 - f'(x_2)F_2 - f'(x_2)f'(x_1)F_1 \end{pmatrix} . \tag{12.8}$$

The next step is to invert the last element in the diagonal, i.e. divide the third row by $1-f'(x_2)f'(x_1)f'(x_3)$. It is clear that if this element is zero at the periodic orbit this step might lead to problems. In many cases this will just mean a slower convergence, but it might throw the Newton iteration completely off. We note that $f'(x_2)f'(x_1)f'(x_3)$ is the stability of the cycle (when the Newton iteration has converged) and that this therefore is not a good method to find marginally stable cycles. We now have

$$\begin{pmatrix} 1 & 0 & -f'(x_3) \\ 0 & 1 & -f'(x_1)f'(x_3) \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix} = \begin{pmatrix} -F_1 \\ -F_2 - f'(x_1)F_1 \\ \frac{-F_3-f'(x_2)F_2-f'(x_2)f'(x_1)F_1}{1-f'(x_2)f'(x_1)f'(x_3)} \end{pmatrix} . \tag{12.9}$$

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. On the left hand side the matrix is now the unit matrix, on the right hand side we have the corrections to our initial guess for the cycle, i.e. we have gone through one step of the Newton iteration scheme.

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, that is 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.

12.1 on p. 288

To illustrate the efficiency of the Newton method we compare it to the inverse iteration method in fig. 12.2. The advantage with respect to speed of Newton's method is obvious.

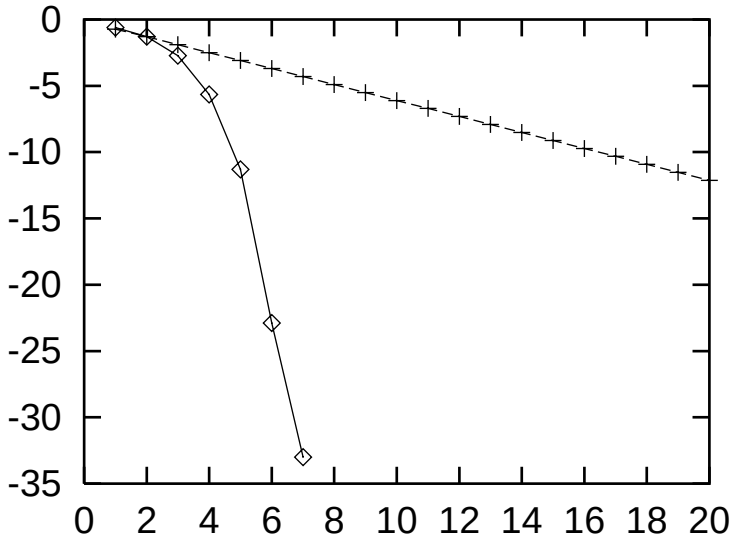


Figure 12.2: Convergence of Newton's method ($\diamondsuit$) vs. inverse iteration ($+$). The error after n iterations searching for the $\overline{01}$-cycle of the logistic map $f(x) = 4x(1-x)$ with an initial starting guess of $x_1 = 0.2, x_2 = 0.8$. y-axis is $\log_{10}$ of the error. The difference between the exponential convergence of the inverse iteration method and the super-exponential convergence of Newton's method is obvious.

12.2 d-dimensional mappings

(F. Christiansen)

Armed with symbolic dynamics informed initial guesses we can utilize the Newton-Raphson iteration in d-dimensions as well.

12.2.1 Newton's method for d-dimensional mappings

Newton's method for 1-dimensional mappings is easily extended to higher dimensions. In this case $f'(x_i)$ is a $[d \times d]$ matrix. $\frac{d}{dx}F(x)$ is then 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. the cycle must not have any marginally stable directions.

Some d-dimensional mappings (such as the Hénon map (3.8)) can be written

as 1-dimensional time delay mappings of the form

$$f(x_i) = f(x_{i-1}, x_{i-2}, \ldots, x_{i-d}). \tag{12.10}$$

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.

12.3 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.

12.3.1 Newton's 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

$$\dot{x} = v(x), \tag{12.11}$$

The corresponding Jacobian matrix $\mathbf{J}$ (4.25) is obtained by integrating the linearized equation (4.31)

$$\dot{\mathbf{J}} = \mathbf{A}\mathbf{J}\,, \qquad A_{ij}(x) = \frac{\partial v_i(x)}{\partial x_j}$$

along the trajectory. The flow and the corresponding Jacobian 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

$$f(x') \approx f(x) + \mathbf{J}(x' - x). \tag{12.12}$$

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 $\mathbf{J}$ 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

$$(1 - \mathbf{J})(x' - x) = -(x - f(x)) = -F(x) \tag{12.13}$$

where $f(x)$ corresponds to integrating from one intersection of the Poincaré surface to another and $\mathbf{J}$ is integrated accordingly. Here we run into problems with the direction along the flow, since this corresponds to a unit eigenvector of $\mathbf{J}$. The matrix $(1 - \mathbf{J})$ does therefore not have full rank. A related problem is that the solution x' of (12.13) 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

$$(x - x_0) \cdot a = 0, \tag{12.14}$$

where a is a vector normal to the Poincaré section and x_0 is any point in the Poincaré section. (12.13) then becomes

$$\begin{pmatrix} 1 - \mathbf{J} & v(x) \\ a & 0 \end{pmatrix} \begin{pmatrix} x' - x \\ \delta T \end{pmatrix} = \begin{pmatrix} -F(x) \\ 0 \end{pmatrix}. \tag{12.15}$$

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

$$\begin{pmatrix} 1-\Lambda & 0 & 0 & 0 \\ 0 & 1-\Lambda_s & 0 & 0 \\ 0 & 0 & 0 & v_z \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \delta_x \\ \delta_y \\ \delta_z \\ \delta t \end{pmatrix} = \begin{pmatrix} -F_x \\ -F_y \\ -F_z \\ 0 \end{pmatrix} . \tag{12.16}$$

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 $\det(\cdot) = v_z \det(1-\mathbf{J}_\perp)$, where $\mathbf{J}_\perp$ is the monodromy matrix for a surface of section transverse to the orbit, see for ex. (22.15).

For periodic orbits (12.15) generalizes in the same way as (12.6), but with n additional equations – one for each point on the Poincaré surface. The Newton setup looks like this

$$\left(\begin{array}{cccccc|ccc} 1 & & & & & -J_n & v_1 & & \\ -J_1 & 1 & & & & & & \ddots & \\ & \cdots & 1 & & & & & & v_n \\ & & \cdots & 1 & & & & & \\ & & & -J_{n-1} & 1 & & & & \\ \hline & a & & & & & 0 & & \\ & & \ddots & & & & & \ddots & \\ & & & & a & & & & 0 \end{array}\right) \begin{pmatrix} \delta_1 \\ \delta_2 \\ \cdot \\ \cdot \\ \delta_n \\ \delta t_1 \\ \cdot \\ \delta t_n \end{pmatrix} = \begin{pmatrix} -F_1 \\ -F_2 \\ \cdot \\ \cdot \\ -F_n \\ 0 \\ \cdot \\ 0 \end{pmatrix} .$$

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 Jacobian 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

$$\begin{pmatrix} 1-\mathbf{J} & v(x) & \nabla H(x) \\ a & 0 & 0 \end{pmatrix} \begin{pmatrix} x'-x \\ \delta t \\ \delta E \end{pmatrix} = \begin{pmatrix} -(x-f(x)) \\ 0 \\ 0 \end{pmatrix} \tag{12.17}$$

simultaneously with

$$H(x') - H(x) = 0. \tag{12.18}$$

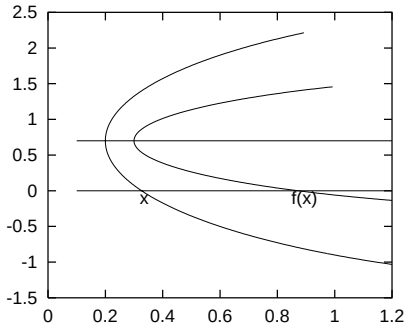


Figure 12.3: Illustration of the optimal Poincaré surface. The original surface $y = 0$ yields a large distance $x - f(x)$ for the Newton iteration. A much better choice is $y = 0.7$.

This last equation is nonlinear. It is often best to treat this separately in the sense that we really solve this equation in each Newton step. This might mean putting in an additional Newton routine to solve the single step of (12.17) and (12.18) together. One might be tempted to linearize (12.18) and put it into (12.17) 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.

12.3.2 Newton's 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 fig. 12.3 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

$$(x' - x) \cdot v(x) = 0. \tag{12.19}$$

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 really 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.

12.4 Periodic orbits as extremal orbits

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 a rough approximation to a longer cycle p of cycle length n_p as a sequence of points $(x_1^{(0)}, x_2^{(0)}, \cdots, x_{n_p}^{(0)})$ with the periodic boundary condition $x_{n_p+1} = x_1$. Suppose you have an iterative method for improving your guess; after k iterations the cost function

$$E(x^{(k)}) = \sum_i^{n_p} \left(x_{i+1}^{(k)} - f(x_i^{(k)}) \right)^2 \tag{12.20}$$

or some other more cleverly constructed function is a measure of the deviation of the kth approximate cycle from the true cycle. This observation motivates variational approaches to determining cycles. We give her two examples of such methods, one for maps and one for billiards. Unlike the Newton-Raphson method, variational methods are very robust. As each step around the cycle is short, they do not suffer from exponential instabilities, and with rather coarse initial guesses one can determine cycles of arbitrary length.

12.4.1 Cyclists relaxation method

(O. Biham 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 (3.8),

$$\begin{aligned} x_{i+1} &= 1 - ax_i^2 + by_i \\ y_{i+1} &= x_i \,, \end{aligned} \tag{12.21}$$

is a stationary solution of the relaxation dynamics defined by the flow

$$\frac{dx_i}{dt} = v_i, \quad i = 1, \ldots, n \tag{12.22}$$

for any vector field $v_i = v_i(x)$ which vanishes on the trajectory. 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 (3.9)

$$v_i = x_{i+1} - 1 + ax_i^2 - bx_{i-1}. \tag{12.23}$$

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)

$$v_i = \frac{d}{dx_i} V_i(x) \,, \qquad V_i(x) = x_i(x_{i+1} - bx_{i-1} - 1) + \frac{a}{3} x_i^3 \,. \tag{12.24}$$

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 (12.22) by

$$\frac{dx_i}{dt} = \sigma_i v_i, \quad i = 1, \ldots, n, \tag{12.25}$$

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 (12.25) - all x_i and $V_i(x)$ change at each integration step - but this is the observation that motivates the method. The differential equations (12.25) then drive an approximate initial guess toward the exact trajectory. A

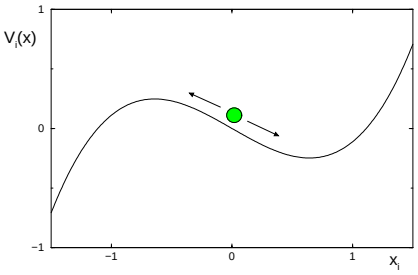


Figure 12.4: "Potential" $V_i(x)$ (12.24) for a typical point along an inital guess trajectory. For $\sigma_i = +1$ the flow is toward the local maximum of $V_i(x)$, and for $\sigma_i = -1$ toward the local minimum. A large deviation of x_i's is needed to destabilize a trajectory passing through such local extremum of $V_i(x)$, hence the basin of attraction is expected to be large.

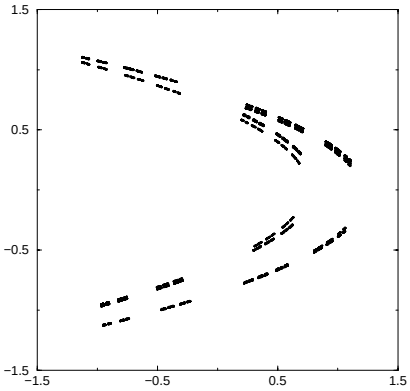


Figure 12.5: The repeller for the Hénon map at $a = 1.8$, $b = 0.3$. (O. Biham)

sketch of the landscape in which x_i converges towards the proper fixed point is given in fig. 12.4. As the "potential" function (12.24) is not bounded for a large $|x_i|$, the flow diverges for initial guesses which are too distant from the true trajectory.

Our aim in this calculation is to find all periodic orbits of period n, 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 condition $x_{n+1} = x_1$. A convenient 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 10.1. The real issue for all searches for periodic orbits, this one included, is how large is the basin of attraction of the desired periodic orbit? There is no easy answer to this question, but empirically it turns out that for the Hénon map such initial guess almost always converges to the desired trajectory as long as the initial $|x|$ is not too large compared to $1/\sqrt{a}$. Fig. 12.4 gives some indication of a typical basin of attraction of the method.

The calculation is carried out by solving the set of n ordinary differential equations (12.25) using a simple Runge-Kutta method with a relatively large step size ($h = 0.1$) until $|v|$ becomes smaller than a given value ε (in a typical calculation $\varepsilon \sim 10^{-7}$). Empirically, in the case that an orbit corresponding to the desired itinerary does not exist, the initial guess escapes to infinity since the "potential" $V_i(x)$ grows without bound.

12.12 on p. 290

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 $n = 1000$. We list all prime cycles up to period 10 for the Hénon map, $a = 1.4$ and $b = 0.3$ are listed in table 12.1. The number of unstable periodic orbits for periods $n \leq 28$ is given in table 12.2. Comparing this with the list of all possible 2-symbol alphabet prime cycles, table 10.1, 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 fig. 12.5. Comparing this set with the strange attractor for the Hénon's parameters fig. 3.4, we note the existence of gaps in the set, cut out by the preimages of the escaping regions.

In practice, this method 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.

12.4.2 Orbit length extremization method for billiards

(Per Dahlquist)

The simplest method for determining billiard cycles 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. 12.2.1 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

$$\partial_i L(s_0 + \delta s) = \partial_i L(s_0) + \sum_j \partial_i \partial_j L(s_0) \delta s_j + \ldots$$

12.9 on p. 289

and use $\mathbf{J}_{ij}(s_0) = \partial_i \partial_j L(s_0)$ in the N-dimensional Newton-Raphson iteration scheme of sect. 12.1.2

$$s_i \mapsto s_i - \sum_j \left(\frac{1}{\mathbf{J}(s)} \right)_{ij} \partial_j L(s) \tag{12.26}$$

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 any of the disks.

12.10 on p. 289

12.11 on p. 290

As an example, the short periods and stabilities of 3-disk cycles computed this way are listed table 12.3.

12.5 Stability of cycles for maps

No matter what method we had used to determine the unstable cycles, the theory to be developed here requires that their stability eigenvalues be evaluated as well. For maps a Jacobian 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 stability matrices according to (4.52). For example, the Jacobian matrix $\mathbf{J}_p$ for a Hénon map (3.8) prime cycle p of length n_p is given by (4.53), and the Jacobian matrix $\mathbf{J}_p$ for a 2-dimensional billiard prime cycle p of length n_p follows from (4.49). As explained on page 81, evaluation of the Jacobian matrix for a flow will require an integration along the prime cycle.

Commentary

Remark **12.1** Intermittency. Intermittency could reduce the efficiency of this method. If only a "small" part of phase space is intermittent then this might work since one needs many of the intermittent cycles in a stability ordered cycle expansion (at least classically). However, if the system is as unbounded as the $(xy)^2$ potential ... forget it ! Sune F. Nielsen

Remark **12.2** Piece-wise linear maps. The Lozi map (3.10) is linear, and 100,000's of cycles can be be easily computed by [2x2] matrix multiplication and inversion.

Remark **12.3** Relaxation method. The relaxation (or gradient) algorithm is one of the methods for solving extremal problems [12]. The method described above was introduced by Biham and Wenzel [13], 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 [14]. The applicability and reliability of the method is discussed in detail by Grassberger, Kantz and Moening [16], who give examples of the ways in which the method fails: (a) it might reach a limit cycle rather than a stationary saddlepoint (that can be remedied by the complex Biham-Wenzel algorithm [14]) (b) different symbol sequences can converge to the same cycle (that is, more refined initial conditions might be needed). Furthermore, Hansen (ref. [17] and chapter 4. of ref. [3]) 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. 12.2.1, 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.

Another method, which is also based on the construction of an artificial dynamics, but of different type, has been introduced by Diakonos and Schmelcher [18]. This method determines cycles ordered by stability, the least unstable cycles being obtained first [20, 19], and is useful in conjunction with the stability ordered cycle expansions that we shall discuss in sect. 13.4.

Remark **12.4** Relation to the Smale horseshoe symbolic dynamics. For a complete horseshoe Hénon repeller (a sufficiently large), such as the one given in fig. 10.17, the signs $\sigma_i \in \{1, -1\}$ are in a 1-to-1 correspondence with the Smale horsheshoe symbolic dynamics $s_i \in \{0, 1\}$:

$$s_i = \begin{cases} 0 & \text{if } \sigma_i = -1\,, \quad x_i < 0 \\ 1 & \text{if } \sigma_i = +1\,, \quad x_i > 0 \end{cases} \quad . \tag{12.27}$$

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 intineraries $\{s_1, s_2, \cdots, s_n\}$ can be much subtler; this is discussed in ref. [16].

Remark **12.5** A compilation of the Hénon map numerical results. For the record - the most accurate estimates of various averages for the Hénon map, Hénon's parameters choice $a = 1.4$, $b = 0.3$, known to the authors, are: the topological entropy (11.1) is $h = 0.4645??$, the Lyapunov exponent $= 0.463$, the Hausdorff dimension $D_H = 1.274(2)$.

References

[12.1] D.W. Moore and E.A. Spiegel, "A thermally excited nonlinear oscillator", *Astrophys. J.*, **143**, 871 (1966).

[12.2] N.H. Baker, D.W. Moore and E.A. Spiegel, *Quar. J. Mech. and Appl. Math.* **24**, 391 (1971).

[12.3] E.A. Spiegel, *Chaos: a mixed metaphor for turbulence*, *Proc. Roy. Soc.* **A413**, 87 (1987).

[12.4] M. Baranger and K.T.R. Davies *Ann. Physics* **177**, 330 (1987).

[12.5] B.D. Mestel and I. Percival, *Physica D* **24**, 172 (1987); Q. Chen, J.D. Meiss and I. Percival, *Physica D* **29**, 143 (1987).

[12.6] find Helleman et all Fourier series methods

[12.7] J.M. Greene, *J. Math. Phys.* **20**, 1183 (1979)

[12.8] H.E. Nusse and J. Yorke, "A procedure for finding numerical trajectories on chaotic saddles" *Physica* **D 36**, 137 (1989).

[12.9] D.P. Lathrop and E.J. Kostelich, "Characterization of an experimental strange attractor by periodic orbits"

[12.10] T. E. Huston, K.T.R. Davies and M. Baranger *Chaos* **2**, 215 (1991).

[12.11] M. Brack, R. K. Bhaduri, J. Law and M. V. N. Murthy, *Phys. Rev. Lett.* **70**, 568 (1993).

[12.12] F. Stummel and K. Hainer, *Praktische Mathematik* (Teubner, Stuttgart 1982).

[12.13] O. Biham and W. Wenzel, *Phys. Rev. Lett.* **63**, 819 (1989).

[12.14] O. Biham and W. Wenzel, *Phys. Rev.* **A 42**, 4639 (1990).

[12.15] P. Grassberger and H. Kantz, *Phys. Lett. A* **113**, 235 (1985).

[12.16] P. Grassberger, H. Kantz and U. Moening, *J. Phys.* **A 43**, 5217 (1989).

[12.17] K.T. Hansen, *Phys. Lett.* **A 165**, 100 (1992).

[12.18] P. Schmelcher and F.K. Diakonos, Phys. Rev. Lett. **78**, 4733 (1997); Phys. Rev. E **57**, 2739 (1998).

[12.19] D. Pingel, P. Schmelcher and F.K. Diakonos, O. Biham, Phys. Rev. E **64**, 026214 (2001).

[12.20] F. K. Diakonos, P. Schmelcher, O. Biham, Phys. Rev. Lett. **81**, 4349 (1998)

[12.21] R.L. Davidchack and Y.C. Lai, Phys. Rev. E **60**, 6172 (1999).

[12.22] 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).

[12.23] F. Moss, *"Chaos under control"*, *Nature* **370**, 615 (1994).

[12.24] J. Glanz, (FIND!), speculated applications of chaos to epilepsy and the brain, chaos-control, *Science* **265**, 1174 (1994).

n	p	(y_p , x_p)	λ_p
1	0	(-1.13135447 , -1.13135447)	1.18167262
	1	(0.63135447 , 0.63135447)	0.65427061
2	01	(0.97580005 , -0.47580005)	0.55098676
4	0111	(-0.70676677 , 0.63819399)	0.53908457
6	010111	(-0.41515894 , 1.07011813)	0.55610982
	011111	(-0.80421990 , 0.44190995)	0.55245341
7	0011101	(-1.04667757 , -0.17877958)	0.40998559
	0011111	(-1.08728604 , -0.28539206)	0.46539757
	0101111	(-0.34267842 , 1.14123046)	0.41283650
	0111111	(-0.88050537 , 0.26827759)	0.51090634
8	00011101	(-1.25487963 , -0.82745422)	0.43876727
	00011111	(-1.25872451 , -0.83714168)	0.43942101
	00111101	(-1.14931330 , -0.48368863)	0.47834615
	00111111	(-1.14078564 , -0.44837319)	0.49353764
	01010111	(-0.52309999 , 0.93830866)	0.54805453
	01011111	(-0.38817041 , 1.09945313)	0.55972495
	01111111	(-0.83680827 , 0.36978609)	0.56236493
9	000111101	(-1.27793296 , -0.90626780)	0.38732115
	000111111	(-1.27771933 , -0.90378859)	0.39621864
	001111101	(-1.10392601 , -0.34524675)	0.51112950
	001111111	(-1.11352304 , -0.36427104)	0.51757012
	010111111	(-0.36894919 , 1.11803210)	0.54264571
	011111111	(-0.85789748 , 0.32147653)	0.56016658
10	0001111101	(-1.26640530 , -0.86684837)	0.47738235
	0001111111	(-1.26782752 , -0.86878943)	0.47745508
	0011111101	(-1.12796804 , -0.41787432)	0.52544529
	0011111111	(-1.12760083 , -0.40742737)	0.53063973
	0101010111	(-0.48815908 , 0.98458725)	0.54989554
	0101011111	(-0.53496022 , 0.92336925)	0.54960607
	0101110111	(-0.42726915 , 1.05695851)	0.54836764
	0101111111	(-0.37947780 , 1.10801373)	0.56915950
	0111011111	(-0.69555680 , 0.66088560)	0.54443884
	0111111111	(-0.84660200 , 0.34750875)	0.57591048
13	1110011101000	(-1.2085766485 , -0.6729999948)	0.19882434
	1110011101001	(-1.0598110494 , -0.2056310390)	0.21072511

Table 12.1: 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 12.4), 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 (Work through exercise **??**). The odds, however, are that it indeed is strange.

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

Table 12.2: 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).

p	Λ_p	T_p
0	9.898979485566	4.000000000000
1	-1.177145519638$\times 10^{1}$	4.267949192431
01	-1.240948019921$\times 10^{2}$	8.316529485168
001	-1.240542557041$\times 10^{3}$	12.321746616182
011	1.449545074956$\times 10^{3}$	12.580807741032
0001	-1.229570686196$\times 10^{4}$	16.322276474382
0011	1.445997591902$\times 10^{4}$	16.585242906081
0111	-1.707901900894$\times 10^{4}$	16.849071859224
00001	-1.217338387051$\times 10^{5}$	20.322330025739
00011	1.432820951544$\times 10^{5}$	20.585689671758
00101	1.539257907420$\times 10^{5}$	20.638238386018
00111	-1.704107155425$\times 10^{5}$	20.853571517227
01011	-1.799019479426$\times 10^{5}$	20.897369388186
01111	2.010247347433$\times 10^{5}$	21.116994322373
000001	-1.205062923819$\times 10^{6}$	24.322335435738
000011	1.418521622814$\times 10^{6}$	24.585734788507
000101	1.525597448217$\times 10^{6}$	24.638760250323
000111	-1.688624934257$\times 10^{6}$	24.854025100071
001011	-1.796354939785$\times 10^{6}$	24.902167001066
001101	-1.796354939785$\times 10^{6}$	24.902167001066
001111	2.005733106218$\times 10^{6}$	25.121488488111
010111	2.119615015369$\times 10^{6}$	25.165628236279
011111	-2.366378254801$\times 10^{6}$	25.384945785676

Table 12.3: All prime cycles up to 6 bounces for the three-disk fundamental domain, center-to-center separation $R = 6$, disk radius $a = 1$. The columns list the cycle itinerary, its expanding eigenvalue Λ_p, and the length of the orbit (if the velocity=1 this is the same as its period or the action). Note that the two 6 cycles $\overline{001011}$ 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)

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

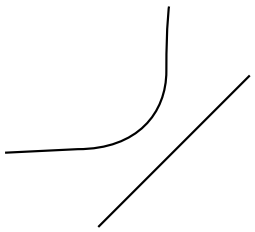
$$f(x) = 4x(1-x)\,. \tag{12.28}$$

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 (4.51) 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 (10.15). This problem is perhaps too hard, but give it a try - the answer is in many introductory books on nolinear 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 3.7 a routine that evaluates the expanding eigenvalue for a given cycle.

12.5 Newton-Raphson method. Implement the Newton-Raphson method in 2-d and apply it to determination of pinball cycles.

12.6 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.

12.7 Cycle stability, helium. Add to the helium integrator of exercise 2.11 a routine that evaluates the expanding eigenvalue for a given cycle.

12.8 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 fig. 23.5.

12.9 Evaluation of 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 12.3. After that check the accuracy of the computed orbits by iterating them forward with your simulator. What is $|f^{T_p}(x)-x|$?

12.10 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, that is some of the minimal length trajectories are blocked by intervening disks.

12.11 Uniqueness of unstable cycles*. Prove that there exists only one 3-disk prime cycle for a given finite admissible prime cycle symbol string. Hints: look at the Poincaré section mappings; can you show that there is exponential contraction to a unique periodic point with a given itinerary? Exercise 12.9 might be helpful in this effort.

12.12 Find cycles of the Hénon map. Apply the method of sect. 12.4.1 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 (11.1), or as the zero of a truncated topological zeta function (11.20). Do your cycles agree with the cycles listed in table 12.1?

12.13 Inverse iteration method for a Hamiltonian repeller. For the complete repeller case (all binary sequences are realized), the cycles are evaluated as follows. According to sect. 3.3, the coordinates of a periodic orbit of length n_p satisfy the equation

$$x_{p,i+1} + x_{p,i-1} = 1 - a x_{p,i}^2 \,, \quad i = 1, ..., n_p \,, \tag{12.29}$$

with the periodic boundary condition $x_{p,0} = x_{p,n_p}$. In the complete repeller case, the Hénon map is a realization of the Smale horseshoe, and the symbolic dynamics has a very simple description in terms of the binary alphabet $\epsilon \in \{0, 1\}$, $\epsilon_{p,i} = (1 + S_{p,i})/2$, where $S_{p,i}$ are the signs of the corresponding cycle point coordinates, $S_{p,i} = \sigma_{x_{p,i}}$. We start with a preassigned sign sequence $S_{p,1}, S_{p,2}, \ldots, S_{p,n_p}$, and a good initial guess for the coordinates $x'_{p,i}$. Using the inverse of the equation (12.29)

$$x''_{p,i} = S_{p,i} \sqrt{\frac{1 - x'_{p,i+1} - x'_{p,i-1}}{a}}, \quad i = 1, ..., n_p \tag{12.30}$$

we converge iteratively, at exponential rate, to the desired cycle points $x_{p,i}$. Given the cycle points, the cycle stabilities and periods are easily computed using (4.53). Verify that the times and the stabilities of the short periodic orbits for the Hénon repeller (3.8) at $a = 6$ are listed in table 12.4; in actual calculations all prime cycles up to topological length $n = 20$ have been computed.

(G. Vattay)

p	Λ_p	$\sum x_{p,i}$
0	$0.71516752438\times10^{1}$	-0.6076252185107
1	$-0.29528463259\times10^{1}$	0.2742918851774
10	$-0.98989794855\times10^{1}$	0.3333333333333
100	$-0.13190727397\times10^{3}$	-0.2060113295833
110	$0.55896964996\times10^{2}$	0.5393446629166
1000	$-0.10443010730\times10^{4}$	-0.8164965809277
1100	$0.57799826989\times10^{4}$	0.0000000000000
1110	$-0.10368832509\times10^{3}$	0.8164965809277
10000	$-0.76065343718\times10^{4}$	-1.4260322065792
11000	$0.44455240007\times10^{4}$	-0.6066540777738
10100	$0.77020248597\times10^{3}$	0.1513755016405
11100	$-0.71068835616\times10^{3}$	0.2484632276044
11010	$-0.58949885284\times10^{3}$	0.8706954728949
11110	$0.39099424812\times10^{3}$	1.0954854155465
100000	$-0.54574527060\times10^{5}$	-2.0341342556665
110000	$0.32222060985\times10^{5}$	-1.2152504370215
101000	$0.51376165109\times10^{4}$	-0.4506624359329
111000	$-0.47846146631\times10^{4}$	-0.3660254037844
110100	$-0.63939998436\times10^{4}$	0.3333333333333
101100	$-0.63939998436\times10^{4}$	0.3333333333333
111100	$0.39019387269\times10^{4}$	0.5485837703548
111010	$0.10949094597\times10^{4}$	1.1514633582661
111110	$-0.10433841694\times10^{4}$	1.3660254037844

Table 12.4: All periodic orbits up to 6 bounces for the Hamiltonian Hénon mapping (12.29) with $a = 6$. Listed are the cycle itinerary, its expanding eigenvalue Λ_p, and its "center of mass". (The last one because we do not understand why the "center of mass" tends to be a simple rational every so often.)

Chapter 13

Cycle expansions

> Recycle... It's the Law!
> Poster, New York City Department of Sanitation

The Euler product representations of spectral determinants (8.9) and dynamical zeta functions (8.12) 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 (8.9), (8.12) 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).

13.1 Pseudocycles and shadowing

How are periodic orbit formulas such as (8.12) evaluated? We start by computing the lengths and stability eigenvalues of the shortest cycles. This always requires numerical work, such as the Newton's 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 tables 12.3 and 12.4. 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 rather slowly).

Expand the dynamical zeta function (8.12) as a formal power series,

$$1/\zeta = \prod_p (1-t_p) = 1 - \sum_{\{p_1 p_2 \cdots p_k\}}{}' (-1)^{k+1} t_{p_1} t_{p_2} \dots t_{p_k} \tag{13.1}$$

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 (13.1) is now compactly written as

$$1/\zeta = 1 - \sum_\pi{}' t_\pi \,. \tag{13.2}$$

For $k > 1$, t_π 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_π.

The pseudocycle weight

$$t_\pi = (-1)^{k+1} \frac{1}{|\Lambda_\pi|} e^{\beta A_\pi - s T_\pi} z^{n_\pi} \,. \tag{13.3}$$

depends on the pseudocycle topological length, integrated observable, period, and stability

$$\begin{aligned} n_\pi &= n_{p_1} + \dots + n_{p_k} \,, \qquad T_\pi = T_{p_1} + \dots + T_{p_k} \\ A_\pi &= A_{p_1} + \dots + A_{p_k} \,, \qquad \Lambda_\pi = \Lambda_{p_1} \Lambda_{p_2} \cdots \Lambda_{p_k} \,. \end{aligned} \tag{13.4}$$

13.1.1 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 (8.12) is given by

$$\begin{aligned} 1/\zeta &= (1-t_0)(1-t_1)(1-t_{01})(1-t_{001})(1-t_{011}) \\ &\quad (1-t_{0001})(1-t_{0011})(1-t_{0111})(1-t_{00001})(1-t_{00011}) \\ &\quad (1-t_{00101})(1-t_{00111})(1-t_{01011})(1-t_{01111}) \dots \end{aligned}$$

(see table 10.1), and the first few terms of the expansion (13.2) ordered by increasing total pseudocycle length are:

$$\begin{aligned} 1/\zeta &= 1 - t_0 - t_1 - t_{01} - t_{001} - t_{011} - t_{0001} - t_{0011} - t_{0111} - \dots \\ &\quad + t_0 t_1 + t_0 t_{01} + t_{01} t_1 + t_0 t_{001} + t_0 t_{011} + t_{001} t_1 + t_{011} t_1 \\ &\quad - t_0 t_{01} t_1 - \dots \end{aligned}$$

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 $\hat{c}_n$. For the binary case this regrouping is given by

$$\begin{aligned} 1/\zeta &= 1 - t_0 - t_1 - [(t_{01} - t_1 t_0)] - [(t_{001} - t_{01} t_0) + (t_{011} - t_{01} t_1)] \\ &\quad - [(t_{0001} - t_0 t_{001}) + (t_{0111} - t_{011} t_1) \\ &\quad + (t_{0011} - t_{001} t_1 - t_0 t_{011} + t_0 t_{01} t_1)] - \dots \\ &= 1 - \sum_f t_f - \sum_n \hat{c}_n \ . \end{aligned} \tag{13.5}$$

All terms in this expansion up to length $n_p = 6$ are given in table 13.1. 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 (see sect. 11.3 and sect. 13.3). 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 (11.20). 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, lies precisely in this organization into nearly cancelling combinations: cycle expansions are dominated by short cycles, with long cycles giving exponentially decaying corrections.

In the case that there is no finite grammar symbolic dynamics to help organize the cycles, the best thing to use is a stability cutoff which we shall discuss in

$-t_0$			
$-t_1$			
$-t_{10}$	$+ t_1t_0$		
$-t_{100}$	$+ t_{10}t_0$		
$-t_{101}$	$+ t_{10}t_1$		
$-t_{1000}$	$+ t_{100}t_0$		
$-t_{1001}$	$+ t_{100}t_1$	$+ t_{101}t_0$	$- t_1t_{10}t_0$
$-t_{1011}$	$+ t_{101}t_1$		
$-t_{10000}$	$+ t_{1000}t_0$		
$-t_{10001}$	$+ t_{1001}t_0$	$+ t_{1000}t_1$	$- t_0t_{100}t_1$
$-t_{10010}$	$+ t_{100}t_{10}$		
$-t_{10101}$	$+ t_{101}t_{10}$		
$-t_{10011}$	$+ t_{1011}t_0$	$+ t_{1001}t_1$	$- t_0t_{101}t_1$
$-t_{10111}$	$+ t_{1011}t_1$		
$-t_{100000}$	$+ t_{10000}t_0$		
$-t_{100001}$	$+ t_{10001}t_0$	$+ t_{10000}t_1$	$- t_0t_{1000}t_1$
$-t_{100010}$	$+ t_{10010}t_0$	$+ t_{1000}t_{10}$	$- t_0t_{100}t_{10}$
$-t_{100011}$	$+ t_{10011}t_0$	$+ t_{10001}t_1$	$- t_0t_{1001}t_1$
$-t_{100101}$	$-t_{100110}$	$+ t_{10010}t_1$	$+ t_{10110}t_0$
	$+ t_{10}t_{1001}$	$+ t_{100}t_{101}$	$- t_0t_{10}t_{101} - t_1t_{10}t_{100}$
$-t_{101110}$	$+ t_{10110}t_1$	$+ t_{1011}t_{10}$	$- t_1t_{101}t_{10}$
$-t_{100111}$	$+ t_{10011}t_1$	$+ t_{10111}t_0$	$- t_0t_{1011}t_1$
$-t_{101111}$	$+ t_{10111}t_1$		

Table 13.1: The binary curvature expansion (13.5) up to length 6, listed in such way that the sum of terms along the pth 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.

sect. 13.4. 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 $\Lambda_{\max}$ larger than the most unstable Λ_p in the data set.

13.1.2 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 β and s. Denote by subscript (i) the ith 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

$$1/\zeta_{(i)} = 1/\zeta_{(i-1)}(1 - t_{(i)}z^{n_{(i)}}) ,$$

and truncating the expansion at each step to a finite polynomial in z^n, $n \leq N$. The result is the Nth order polynomial approximation

$$1/\zeta_N = 1 - \sum_{n=1}^{N} \hat{c}_n z^n \,. \tag{13.6}$$

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 (13.6) 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 (13.6) as function of s.

13.1.3 Evaluation of traces, spectral determinants

Due to the lack of factorization of the full pseudocycle weight, $\det(\mathbf{1} - \mathbf{J}_{p_1 p_2}) \neq \det(\mathbf{1} - \mathbf{J}_{p_1}) \det(\mathbf{1} - \mathbf{J}_{p_2})$, the cycle expansions for the spectral determinant (8.9) 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 (7.9) truncated to all prime cycles p and their repeats such that $n_p r \leq N$:

$$\begin{aligned}
\operatorname{tr} \frac{z\mathcal{L}}{1 - z\mathcal{L}}\Big|_{(i)} &= \operatorname{tr} \frac{z\mathcal{L}}{1 - z\mathcal{L}}\Big|_{(i-1)} + n_{(i)} \sum_{r=1}^{n_{(i)} r \leq N} \frac{e^{(\beta \cdot A_{(i)} - s T_{(i)}) r}}{\left| \prod \left(1 - \Lambda_{(i),j}^{r}\right) \right|} z^{n_{(i)} r} \\
\operatorname{tr} \frac{z\mathcal{L}}{1 - z\mathcal{L}}\Big|_{N} &= \sum_{n=1}^{N} C_n z^n \,, \qquad C_n = \operatorname{tr} \mathcal{L}^n \,.
\end{aligned} \tag{13.7}$$

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 (β, s) parameter values. Now that we have an expansion for the trace formula (7.8) as a power

series, we compute the Nth order approximation to the spectral determinant (8.3)

$$\det(1-z\mathcal{L})|_N = 1-\sum_{n=1}^{N} Q_n z^n\,, \qquad Q_n = Q_n(\mathcal{L}) = n\text{th cumulant} \tag{13.8}$$

as follows. The logarithmic derivative relation (8.4) yields

$$\begin{aligned}\left(\mathrm{tr}\,\frac{z\mathcal{L}}{1-z\mathcal{L}}\right)\det(1-z\mathcal{L}) &= -z\frac{d}{dz}\det(1-z\mathcal{L})\\ (C_1 z + C_2 z^2 + \cdots)(1-Q_1 z - Q_2 z^2 - \cdots) &= Q_1 z + 2Q_2 z^2 + 3Q_3 z^3 \cdots\end{aligned}$$

so the nth order term of the spectral determinant cycle (or in this case, the cumulant) expansion is given recursively by the trace formula expansion coefficients

$$Q_n = \frac{1}{n}\left(C_n - C_{n-1}Q_1 - \cdots C_1 Q_{n-1}\right)\,. \tag{13.9}$$

Given the trace formula (13.7) truncated to z^N we now also have the spectral determinant truncated to z^N.

The same method can also be used to compute the dynamical zeta function cycle expansion (13.6), by replacing $\prod\left(1-\Lambda_{(i),j}^r\right)$ in (13.7) by the product of expanding eigenvalues $\Lambda_{(i)} = \prod_e \Lambda_{(i),e}$, as in sect. 8.3.

The calculation of the leading eigenvalue of a given 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 (13.8) is the cycle expansion for (8.6), the spectral determinant $\det(s-\mathcal{A})$. We vary s in cycle weights, and determine the eigenvalue s_α by finding $s = s_\alpha$ for which (13.8) vanishes. The convergence of a leading eigenvalue for a nice hyperbolic system is illustrated by the listing of pinball escape rate γ estimates computed from truncations of (13.5) and (13.8) to different maximal cycle lengths, table 13.2.

The pleasant surprise is that the coefficients in these expansions can be proven to fall off exponentially or even faster fast track: sect. 9, p. 169, 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.

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	0.41033840776934648 2	0.4103383	0.410338
	7	0.4103384077693464892		0.4103396
	8	0.41033840776934648933846 8		
	9	0.41033840776934648933846 13074		
	10	0.41033840776934648933846 13078192		
	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		

Table 13.2: 3-disk repeller escape rates computed from the cycle expansions of the spectral determinant (8.6) and the dynamical zeta function (8.12), 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 (13.5), and the fifth from the full 3-disk cycle expansion (13.31). 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.)

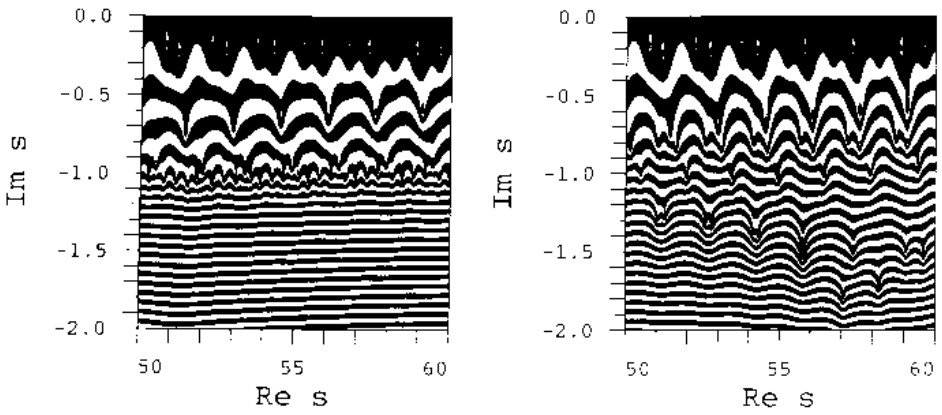


Figure 13.1: Examples of the complex s plane scans: contour plots of the logarithm of the absolute values of (a) $1/\zeta(s)$, (b) spectral determinant $\det(s-\mathcal{A})$ for the 3-disk system, separation $a:\ R=6$, A_1 subspace are evaluated numerically. The eigenvalues of the evolution operator $\mathcal{L}$ 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)

13.1.4 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 $\ln|\det(1-\mathcal{L})|$ or the phase of spectral determinants and dynamical zeta functions as functions of complex 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, fig. 13.1, 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 fig. 13.1), a zero $1/\zeta(s)=0$ can now be easily determined by standard numerical methods, such as the iterative Newton algorithm (12.3)

$$s_{n+1} = s_n - \left(\zeta(s_n)\frac{\partial}{\partial s}\zeta^{-1}(s_n)\right)^{-1} = s_n - \frac{1/\zeta(s_n)}{\langle T\rangle_\zeta}\ . \tag{13.10}$$

The derivative of $1/\zeta(s)$ required for the Newton iteration is given by the cycle expansion (13.18) that we need to evaluate anyhow, as $\langle T\rangle_\zeta$ enters our cycle

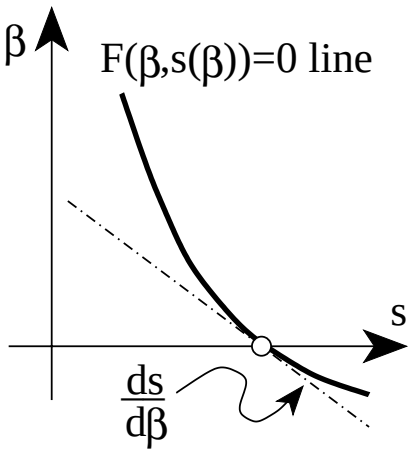


Figure 13.2: The eigenvalue condition is satisfied on the curve $F = 0$ the (β, s) plane. The expectation value of the observable (6.12) is given by the slope of the curve.

averaging formulas.

13.2 Cycle formulas for dynamical averages

The eigenvalue condition in any of the three forms that we have given so far - the level sum (14.18), the dynamical zeta function (13.2), the spectral determinant (13.8):

$$1 = \sum_i^{(n)} t_i\,, \qquad t_i = t_i(\beta, s(\beta)) = \frac{1}{|\Lambda_i|} e^{\beta\cdot A_i - s(\beta)T_i} \tag{13.11}$$

$$0 = 1 - \sum_\pi{}' t_\pi\,, \qquad t_\pi = t_\pi(z, \beta, s(\beta)) \tag{13.12}$$

$$0 = 1 - \sum_{n=1}^{\infty} Q_n\,, \qquad Q_n = Q_n(\beta, s(\beta))\,, \tag{13.13}$$

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 β is sketched in fig. 13.2; 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 by taking derivatives of the eigenvalue condition. Evaluated along $F = 0$, the first derivative leads to

$$\begin{aligned} 0 &= \frac{d}{d\beta} F(\beta, s(\beta)) \\ &= \frac{\partial F}{\partial \beta} + \frac{ds}{d\beta}\frac{\partial F}{\partial s}\bigg|_{s=s(\beta)} \quad \Longrightarrow \quad \frac{ds}{d\beta} = -\frac{\partial F}{\partial \beta}\Big/\frac{\partial F}{\partial s}\,, \end{aligned} \tag{13.14}$$

and the second derivative of $F(\beta, s(\beta)) = 0$ yields

$$\frac{d^2 s}{d\beta^2} = -\left[\frac{\partial^2 F}{\partial \beta^2} + 2\frac{ds}{d\beta}\frac{\partial^2 F}{\partial \beta \partial s} + \left(\frac{ds}{d\beta}\right)^2 \frac{\partial^2 F}{\partial s^2}\right] / \frac{\partial F}{\partial s}\,. \tag{13.15}$$

Denoting by

$$\begin{aligned} \langle A\rangle_F &= -\left.\frac{\partial F}{\partial \beta}\right|_{\beta, s=s(\beta)}\,, \qquad \langle T\rangle_F = \left.\frac{\partial F}{\partial s}\right|_{\beta, s=s(\beta)} \\ \left\langle (A-\langle A\rangle)^2\right\rangle_F &= \left.\frac{\partial^2 F}{\partial \beta^2}\right|_{\beta, s=s(\beta)} \end{aligned} \tag{13.16}$$

respectively the mean cycle expectation value of A and the mean cycle period computed from the $F(\beta, s(\beta)) = 0$ condition, we obtain the cycle averaging formulas for the expectation value of the observable (6.12) and its variance

$$\begin{aligned} \langle a\rangle &= \frac{\langle A\rangle_F}{\langle T\rangle_F} \\ \left\langle (a-\langle a\rangle)^2\right\rangle &= \frac{1}{\langle T\rangle_F}\left\langle (A-\langle A\rangle)^2\right\rangle_F\,. \end{aligned} \tag{13.17}$$

These formulas are the central result of the periodic orbit theory. As we shall see below, for each choice of the eigenvalue condition function $F(\beta, s)$ in (14.18), (13.2) and (13.8), the above quantities have explicit cycle expansions.

13.2.1 Dynamical zeta function cycle expansions

For the dynamical zeta function condition (13.12), the cycle averaging formulas (13.14), (13.17) require evaluation of the derivatives of dynamical zeta function at a given eigenvalue. Substituting the cycle expansion (13.2) for dynamical zeta function we obtain

$$\begin{aligned} \langle A\rangle_\zeta &:= -\frac{\partial}{\partial \beta}\frac{1}{\zeta} = {\sum}' A_\pi t_\pi \\ \langle T\rangle_\zeta &:= \frac{\partial}{\partial s}\frac{1}{\zeta} = {\sum}' T_\pi t_\pi\,, \quad \langle n\rangle_\zeta := -z\frac{\partial}{\partial z}\frac{1}{\zeta} = {\sum}' n_\pi t_\pi\,, \end{aligned} \tag{13.18}$$

where the subscript in $\langle\cdots\rangle_\zeta$ stands for the dynamical zeta function average over prime cycles, A_π, T_π, and n_π are evaluated on pseudocycles (13.4), and

pseudocycle weights $t_\pi = t_\pi(z, \beta, s(\beta))$ are evaluated at the eigenvalue $s(\beta)$. In most applications, $s(\beta)$ is typically the leading eigenvalue.

For bounded flows the leading eigenvalue (the escape rate) vanishes, $s(0) = 0$, so

$$\langle A\rangle_\zeta = \sum_\pi{}' (-1)^{k+1} \frac{A_{p_1} + A_{p_2} \cdots + A_{p_k}}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}\,, \tag{13.19}$$

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

$$\begin{aligned}\langle T\rangle_\zeta &= \frac{T_0}{|\Lambda_0|} + \frac{T_1}{|\Lambda_1|} + \left(\frac{T_{01}}{|\Lambda_{01}|} - \frac{T_0 + T_1}{|\Lambda_0\Lambda_1|}\right) \\ &\quad + \left(\frac{T_{001}}{|\Lambda_{001}|} - \frac{T_{01} + T_0}{|\Lambda_{01}\Lambda_0|}\right) + \left(\frac{T_{011}}{|\Lambda_{011}|} - \frac{T_{01} + T_1}{|\Lambda_{01}\Lambda_1|}\right) + \ldots\,. \end{aligned} \tag{13.20}$$

Note that the cycle expansions for averages are grouped into the same shadowing combinations as the dynamical zeta function cycle expansion (13.5), 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 (13.17). Assuming zero mean drift $\langle a\rangle = 0$, the cycle expansion for the variance $\left\langle (A - \langle A\rangle)^2\right\rangle_\zeta$ is given by

$$\left\langle A^2\right\rangle_\zeta = \sum{}' (-1)^{k+1} \frac{(A_{p_1} + A_{p_2} \cdots + A_{p_k})^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}\,. \tag{13.21}$$

13.2.2 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 13.2. 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 9, for such systems the spectral determinant constructed from the same cycle data base is entire, and its cycle expansion converges faster than exponentially. Hence in practice, the best convergence is attained by the spectral determinant cycle expansion (13.13) 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 (13.9) and

(13.7). The cycle averaging formulas formulas are exact, and highly convergent for nice hyperbolic dynamical systems. We shall illustrate the utility of such cycle expansions in chapter **??**.

13.2.3 Continuous vs. discrete mean return time

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

$$\langle a\rangle_{\rm dscr} = \langle a\rangle \langle T\rangle_\zeta / \langle n\rangle_\zeta \ , \tag{13.22}$$

where $\langle n\rangle_\zeta$ is the average of the number of bounces n_p along the cycle p.

13.3 Cycle expansions for finite alphabets

A finite Markov graph like the one given in fig. 10.15(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 (11.16) 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:

$$\det(1-T) = 1 - t_0 - t_{0011} - t_{0001} - t_{00011} + t_0 t_{0011} + t_{0011} t_{0001} \tag{13.23}$$

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 (11.17).

The determinant (13.23) is exact for the finite graph fig. 10.15(e), as well as for the associated transfer operator of sect. 5.2.1. For the associated (infinite dimensional) evolution operator, it is the beginning of the cycle expansion of the corresponding dynamical zeta function:

$$\begin{aligned} 1/\zeta \ &= \ 1 - t_0 - t_{0011} - t_{0001} + t_{0001}t_{0011} \\ &\qquad -(t_{00011} - t_0 t_{0011} + \ldots \text{curvatures}) \ldots \end{aligned} \tag{13.24}$$

The cycles 0, 0001 and 0011 are the *fundamental* cycles introduced in (13.5); they are not shadowed by any combinations of shorter cycles, and are the basic building blocks of the dynamics generated by iterating the pruning rules (10.28). All other cycles appear together with their shadows (for example, $t_{00011} - t_0 t_{0011}$ combination is of that type) 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 (13.2) 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 (11.20) 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 10.4 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, that is, only for n larger than the lengths of the fundamental cycles are the curvatures $\hat{c}_n$, a measure of the deviations between long orbits and their short cycle approximants, expected to fall off rapidly with n.

13.4 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 16, 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

$$|\Lambda_{p_1} \cdots \Lambda_{p_k}| \leq \Lambda_{\max} \tag{13.25}$$

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.

13.4.1 Stability ordering for bad grammars

For generic flows it is often not clear what partition of the phase 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 (13.5) 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 finding cycles preferentially finds the least unstable cycles, regardless of their topological length. Another practical advantage of the method (in contrast to the Newton method searches) is that it only finds cycles in a given connected ergodic component of phase space, even if isolated cycles or other ergodic regions exist elsewhere in the phase space.

Why should stability ordered cycle expansion of a dynamical zeta function converge better than the rude trace formula (14.9)? The argument has essentially already been laid out in sect. 11.7: in truncations that respect shadowing most of the pseudocycles appear in shadowning combinations and nearly cancel, and only the relatively small subset affected by the longer and longer pruning rules appears not shadowed. So the error is typically of the order of $1/\Lambda$, smaller by factor e^{hT} than the trace formula (14.9) error, where h is the entropy and T typical cycle length for cycles of stability Λ.

13.4.2 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 $1/\Lambda$ weigth 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 Λ 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 should always be tested for robustness by varying the cutoff. If this introduces significant variations, smoothing is probably necessary.

13.4.3 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 1, the simplest of which is the Farey map

$$f(x) = \begin{cases} x/(1-x) & 0 \le x \le 1/2 \quad L \\ (1-x)/x & 1/2 \le x \le 1 \quad R\,, \end{cases} \tag{13.26}$$

a map which will reappear in chapter 19 in the the study of circle maps.

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 $x_L = 0$, for which the stability

equals $\Lambda_L = 1$. This fixed point does not participate directly in the dynamics and is omitted from cycle expansions. Its presence is felt in the stabilities of neighboring cycles with n consecutive repeats of the symbol L's whose stability falls of only as $\Lambda \sim n^2$, in contrast to the most unstable cycles with n consecutive R's which are exponentially unstable, $|\Lambda_{LR^n}| \sim [(\sqrt{5}+1)/2]^{2n}$.

The symbolic dynamics is of complete binary type, so a quick count in the style of sect. 11.5.2 leads to a total of 74,248,450 prime cycles of length 30 or less, not including the marginal point $x_L = 0$. Evaluating a cycle expansion to this order would be no mean computational feat. However, the least unstable cycle omitted has stability of roughly $\Lambda_{RL^{30}} \sim 30^2 = 900$, and so amounts to a 0.1% correction. The situation may be much worse than this estimate suggests, because the next, RL^{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 (see fig. 13.3).

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 fig. 13.3. We see that topological length truncation schemes are hopelessly bad in this case; stability length truncations are somewhat better, but still rather bad. As we shall show in sect. **??**, 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.

13.5 Dirichlet series

A Dirichlet series is defined as

$$f(s) = \sum_{j=1}^{\infty} a_j e^{-\lambda_j s} \tag{13.27}$$

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 (13.2) ordered by the increasing pseudocycle periods are often Dirichlet series. For example, for the

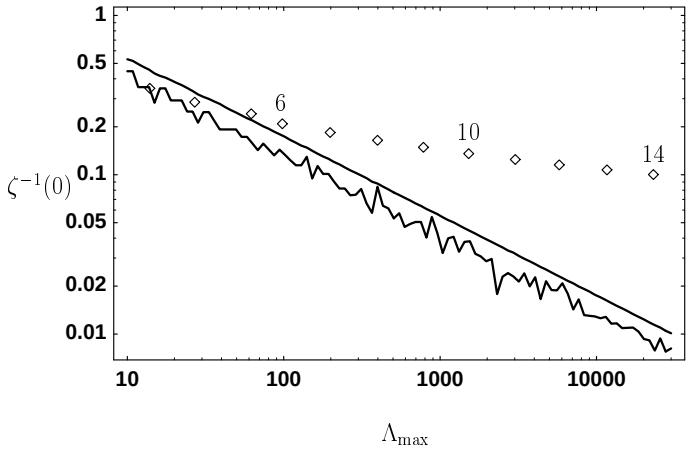


Figure 13.3: Comparison of cycle expansion truncation schemes for the Farey map (13.26); the deviation of the truncated cycles expansion for $|1/\zeta_N(0)|$ from the exact flow conservation value $1/\zeta(0) = 0$ is a measure of the accuracy of the truncation. The jagged line is logarithm of the stability ordering truncation error; the smooth line is smoothed according to sect. 13.4.2; 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.

pseudocycle weight (13.3), 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

$$a_\pi = \frac{e^{\beta\cdot(A_{p_1}+A_{p_2}+\ldots+A_{p_k})}}{|\Lambda_{p_1}\Lambda_{p_2}\ldots\Lambda_{p_k}|} d_\pi \,,$$

where d_π is a degeneracy factor, in the case that d_π pseudocycles have the same weight.

If the series $\sum |a_j|$ diverges, the Dirichlet series is absolutely convergent for $\mathrm{Re}\, s > \sigma_a$ and conditionally convergent for $\mathrm{Re}\, s > \sigma_c$, where σ_a is the *abscissa of absolute convergence*

$$\sigma_a = \lim_{N\to\infty} \sup \frac{1}{\lambda_N} \ln \sum_{j=1}^{N} |a_j| \,, \tag{13.28}$$

and σ_c is the *abscissa of conditional convergence*

$$\sigma_c = \lim_{N\to\infty} \sup \frac{1}{\lambda_N} \ln \left| \sum_{j=1}^{N} a_j \right| \,. \tag{13.29}$$

We shall encounter another example of a Dirichlet series in the semiclassical quantization chapter **??**, where the inverse Planck constant is a complex variable $s = i/\hbar$, $\lambda_\pi = S_{p_1} + S_{p_2} + \ldots + S_{p_k}$ is the pseudocycle action, and $a_\pi = 1/\sqrt{|\Lambda_{p_1}\Lambda_{p_2}\ldots\Lambda_{p_k}|}$ (times possible degeneracy and topological phase factors). As the action is in general not a linear function of energy (except for billiards and for scaling potentials, where a variable s can be extracted from S_p), semiclassical cycle expansions are Dirichlet series in variable $s = i/\hbar$ but not in E, the complex energy variable.

Commentary

Remark **13.1** Pseudocycle expansions. Bowen's introduction of shadowing ϵ-pseudoorbits [13] was a significant contribution to Smale's theory. Expression "pseudoorbits" seems to have been introduced in the Parry and Pollicott's 1983 paper [5]. Following them M. Berry [8] had used the expression "pseudoorbits" in his 1986 paper on Riemann zeta and quantum chaology. Cycle and curvature expansions of dynamical zeta functions and spectral determinants were introduced in refs. [9, 1]. Some literature [**?**] refers to the pseudoorbits as "composite orbits", and to the cycle expansions as "Dirichlet series" (see also remark 13.6 and sect. 13.5).

Remark **13.2** Cumulant expansion. To statistical mechanician the curvature expansions are very reminiscent of cumulant expansions. Indeed, (13.9) is the standard Plemelj-Smithies cumulant formula (J.25) for the Fredholm determinant, discussed in more detail in appendix J.

Remark **13.3** 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 17) reduces the number of nth level terms by another factor. While the formulation of the theory from the trace (7.24) to the cycle expansion (13.5) 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 **13.4** Shadowing cycle-by-cycle. A glance at the low order curvatures in the table 13.1 leads to a 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. [2], table 1). However, splitting $\hat{c}_n$ into individual cycle curvatures is not possible in general [**?**]; the first example of such ambiguity in the binary cycle expansion is given by the

t_{001011}, t_{010011} $0 \leftrightarrow 1$ symmetric pair of 6-cycles; the counterterm $t_{001}t_{011}$ in table 13.1 is shared by the two cycles.

Remark **13.5** Stability ordering. The stability ordering was introduced by Dahlqvist and Russberg [11] 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 [12] for the Lorentz gas which is hyperbolic but the symbolic dynamics is highly pruned, and Dettmann and Cvitanović [13] 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.

Remark **13.6** Are cycle expansions Dirichlet series? Even though some literature [**?**] 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 have much to do with the actual convergence of the theory.

Résumé

A *cycle expansion* is a series representation of a dynamical zeta function, trace formula or a spectral determinant, with products in (8.12), (22.13) expanded as sums over *pseudocycles*, products of the prime cycle weigths t_p.

If a flow is hyperbolic and has a topology of a Smale horseshoe, the associated zeta functions have nice analytic structure: 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.

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 open disks billiards, 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

$$\begin{aligned} \langle a \rangle &= \langle A \rangle_\zeta \,/\, \langle T \rangle_\zeta \\ \langle A \rangle_\zeta &= -\frac{\partial}{\partial \beta}\frac{1}{\zeta} = \sum{}' A_\pi t_\pi \,, \qquad \langle T \rangle_\zeta = \frac{\partial}{\partial s}\frac{1}{\zeta} = \sum{}' T_\pi t_\pi \end{aligned}$$

yield the expectation value (the chaotic, ergodic average over the non–wandering set) of the observable $a(x)$.

References

[13.1] R. Artuso, E. Aurell and P. Cvitanović, "Recycling of strange sets I: Cycle expansions", *Nonlinearity* **3**, 325 (1990).

[13.2] R. Artuso, E. Aurell and P. Cvitanović, "Recycling of strange sets II: Applications", *Nonlinearity* **3**, 361 (1990).

[13.3] S. Grossmann and S. Thomae, *Z. Naturforsch.* **32 a**, 1353 (1977); reprinted in ref. [4].

[13.4] *Universality in Chaos, 2. edition*, P. Cvitanović, ed., (Adam Hilger, Bristol 1989).

[13.5] F. Christiansen, P. Cvitanović and H.H. Rugh, *J. Phys* **A 23**, L713 (1990).

[13.6] J. Plemelj, *"Zur Theorie der Fredholmschen Funktionalgleichung"*, *Monat. Math. Phys.* **15**, 93 (1909).

[13.7] F. Smithies, *"The Fredholm theory of integral equations"*, *Duke Math.* **8**, 107 (1941).

[13.8] 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).

[13.9] P. Cvitanović, *"Invariant measurements of strange sets in terms of cycles"*, *Phys. Rev. Lett.* **61**, 2729 (1988).

[13.10] B. Eckhardt and G. Russberg, *Phys. Rev.* **E 47**, 1578 (1993).

[13.11] 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).

[13.12] C. P. Dettmann and G. P. Morriss, *Phys. Rev. Lett.* **78**, 4201 (1997).

[13.13] C. P. Dettmann and P. Cvitanović, *Cycle expansions for intermittent diffusion Phys. Rev. E* **56**, 6687 (1997); chao-dyn/9708011.

Exercises

13.1 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.

13.2 Escape rate for a 1-d repeller. (Continuation of exercise 8.1 - easy, but long)
Consider again the quadratic map (8.31)

$$f(x) = Ax(1-x)$$

on the unit interval, for definitivness 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. 12.1.1).

(c) Determine the leading zero of the zeta function (8.12) using the weigths $t_p = z^{n_p}/|\Lambda_p|$ where Λ_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 13.2. Compare the coefficients of the spectral determinant and the zeta function cycle expansions. Which expansion converges faster?

(Per Rosenqvist)

13.3 Escape rate for the Ulam map. Check that the escape rate for the Ulam map, $A = 4$ in (8.31)

$$f(x) = 4x(1-x),$$

equals zero. 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.

13.4 Pinball escape rate, semi-analytical. Estimate the 3-disk pinball escape rate for $R : a = 6$ by substituting analytical cycle stabilities and periods (exercise 4.4 and exercise 4.5) into the appropriate binary cycle expansion. Compare with the numerical estimate exercise 8.11

13.5 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 12.6 into the binary cycle expansion.

13.6 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 (13.5) 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?

13.7 Counting the 3-disk pinball counterterms. Verify that the number of terms in the 3-disk pinball curvature expansion (13.30) is given by

$$\begin{aligned}\prod_p (1+t_p) &= \frac{1-3z^4-2z^6}{1-3z^2-2z^3} = 1+3z^2+2z^3+\frac{z^4(6+12z+2z^2)}{1-3z^2-2z^3} \\ &= 1+3z^2+2z^3+6z^4+12z^5+20z^6+48z^7+84z^8+184z^9+\ldots\end{aligned}$$

This means that, for example, c_6 has a total of 20 terms, in agreement with the explicit 3-disk cycle expansion (13.31).

13.8 3–disk unfactorized zeta cycle expansions. Check that the curvature expansion (13.2) for the 3-disk pinball, assuming no symmetries between disks, is given by

$$\begin{aligned} 1/\zeta &= (1-z^2t_{12})(1-z^2t_{13})(1-z^2t_{23})(1-z^3t_{123})(1-z^3t_{132}) \\ &\quad (1-z^4t_{1213})(1-z^4t_{1232})(1-z^4t_{1323})(1-z^5t_{12123})\cdots \\ &= 1-z^2t_{12}-z^2t_{23}-z^2t_{31}-z^3t_{123}-z^3t_{132} \\ &\quad -z^4[(t_{1213}-t_{12}t_{13})+(t_{1232}-t_{12}t_{23})+(t_{1323}-t_{13}t_{23})] \\ &\quad -z^5[(t_{12123}-t_{12}t_{123})+\cdots]-\cdots \end{aligned} \tag{13.30}$$

The symmetrically arranged 3-disk pinball cycle expansion of the Euler product (13.2) (see table 11.4 and fig. 17.2) is given by:

$$\begin{aligned} 1/\zeta &= (1-z^2t_{12})^3(1-z^3t_{123})^2(1-z^4t_{1213})^3 \\ &\quad (1-z^5t_{12123})^6(1-z^6t_{121213})^6(1-z^6t_{121323})^3\ldots \\ &= 1-3z^2\,t_{12}-2z^3\,t_{123}-3z^4\,(t_{1213}-t_{12}^2)-6z^5\,(t_{12123}-t_{12}t_{123}) \\ &\quad -z^6\,(6\,t_{121213}+3\,t_{121323}+t_{12}^3-9\,t_{12}t_{1213}-t_{123}^2) \\ &\quad -6z^7\,(t_{1212123}+t_{1212313}+t_{1213123}+t_{12}^2t_{123}-3\,t_{12}t_{12123}-t_{123}t_{1213}) \\ &\quad -3z^8\,(2\,t_{12121213}+t_{12121313}+2\,t_{12121323}+2\,t_{12123123} \\ &\qquad +2\,t_{12123213}+t_{12132123}+3\,t_{12}^2t_{1213}+t_{12}t_{123}^2 \\ &\qquad -6\,t_{12}t_{121213}-3\,t_{12}t_{121323}-4\,t_{123}t_{12123}-t_{1213}^2)-\cdots \end{aligned} \tag{13.31}$$

Remark **13.7** Unsymmetrized cycle expansions. The above 3-disk cycle expansions might be useful for cross-checking purposes, but, as we shall see in chapter 17, they are not recommended for actual computations, as the factorized zeta functions yield much better convergence.

13.9 4–disk unfactorized dynamical zeta function cycle expansions For the symmetricly 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 11.2):

$$\begin{aligned} 1/\zeta &= (1-z^2t_{12})^4(1-z^2t_{13})^2(1-z^3t_{123})^8(1-z^4t_{1213})^8(1-z^4t_{1214})^4 \\ &\quad (1-z^4t_{1234})^2(1-z^4t_{1243})^4(1-z^5t_{12123})^8(1-z^5t_{12124})^8(1-z^5t_{12134})^8 \\ &\quad (1-z^5t_{12143})^8(1-z^5t_{12313})^8(1-z^5t_{12413})^8\cdots \end{aligned} \tag{13.32}$$

and the cycle expansion is given by

$$\begin{aligned} 1/\zeta \;=\; & 1 - z^2(4\,t_{12} + 2\,t_{13}) - 8z^3\,t_{123} \\ & -z^4(8\,t_{1213} + 4\,t_{1214} + 2\,t_{1234} + 4\,t_{1243} - 6\,t_{12}^2 - t_{13}^2 - 8\,t_{12}t_{13}) \\ & -8z^5(t_{12123} + t_{12124} + t_{12134} + t_{12143} + t_{12313} + t_{12413} - 4\,t_{12}t_{123} - 2\,t_{13}t_{123}) \\ & -4z^6(2\,S_8 + \,S_4 + \,t_{12}^3 + 3\,t_{12}^2\,t_{13} + \,t_{12}t_{13}^2 - 8\,t_{12}t_{1213} - 4\,t_{12}t_{1214} \\ & -2\,t_{12}t_{1234} - 4\,t_{12}t_{1243} - 4\,t_{13}t_{1213} - 2\,t_{13}t_{1214} - \,t_{13}t_{1234} \\ & -2\,t_{13}t_{1243} - 7\,t_{123}^2) - \cdots \end{aligned} \tag{13.33}$$

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 11.4.

13.10 Tail resummations. A simple illustration of such tail resummation is the ζ function for the Ulam map (12.28) 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 ζ function by a simple trick: we note that if the value of any n-cycle eigenvalue were t^n, (8.18) 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 ζ function

$$1/\zeta(z) \;=\; \frac{(1-2t)(1-t_0)}{(1-t)} \tag{13.34}$$

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 (13.5) unexpanded, and noted that all curvature combinations in (13.5) which include the t_0 factor are unbalanced, so that the cycle expansion is an infinite series:

$$\prod_p (1 - t_p) \;=\; (1 - t_0)(1 - t - t^2 - t^3 - t^4 - \ldots) \tag{13.35}$$

(we shall return to such infinite series in chapter 16). The geometric series in the brackets sums up to (13.34). 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 ζ function (13.34).

Chapter 14

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 lan 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. We start out by explaining how qualitatively how local exponential instability and exponential growth in topologically distinct trajectories lead to a global exponential instability.

14.1 Escape rates

We start by verifying the claim (6.11) that for a nice hyperbolic flow the trace of the evolution operator grows exponentially with time. Consider again the game of pinball of fig. 1.1. Designate by $\mathcal{M}$ a phase 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 phase space region $\mathcal{M}$ and recur within that region at the time t is given by

$$\hat{\Gamma}_{\mathcal{M}}(t) = \frac{1}{|\mathcal{M}|}\int\int_{\mathcal{M}} dxdy\,\delta\big(y - f^t(x)\big)\ . \tag{14.1}$$

This quantity is eminently measurable and physically interesting in a variety of problems spanning from 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 phase 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 fig. 1.1 is gone for good.

These observations can be made more concrete by examining the pinball phase space of fig. 1.7. With each pinball bounce the initial conditions that survive get thinned out, each strip yielding two thiner strips within it. The total fraction of survivors (1.2) after n bounces is given by

$$\hat{\Gamma}_n = \frac{1}{|\mathcal{M}|} \sum_i^{(n)} |\mathcal{M}_i| \,, \tag{14.2}$$

where i is a binary label of the ith strip, and $|\mathcal{M}_i|$ is the area of the ith 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 ith strip, assume that the spreading of a ray of trajectories per bounce is given by a factor Λ, the mean value of the expanding eigenvalue of the corresponding Jacobian 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. 89) is right in spirit, but not without drawbacks. One problem is that in general the eigenvalues of a Jacobian matrix have no invariant meaning; they depend on the choice of coordinates. However, we saw in chapter 7 that the sizes of neighborhoods are determined by stability eigenvalues of periodic points, and those are invariant under smooth coordinate transformations.

In this approximation $\hat{\Gamma}_n$ receives 2^n contributions of equal size

$$\hat{\Gamma}_1 \sim \frac{1}{\Lambda} + \frac{1}{\Lambda}, \cdots \quad , \hat{\Gamma}_n \sim \frac{2^n}{\Lambda^n} = e^{-n(\lambda - h)} := e^{-n\gamma} \,, \tag{14.3}$$

up to preexponential factors. We see here the interplay of the two key ingredients of chaos first alluded to in sect. 1.3.1: the escape rate γ 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 we shall see in (15.16), with correctly defined "entropy" this result is exact.

As at each bounce one loses routinely the same fraction of trajectories, one expects the sum (14.2) to fall off exponentially with n. More precisely, by the hyperbolicity assumption of sect. 7.1.1 the expanding eigenvalue of the Jacobian matrix of the flow is exponentially bounded from both above and below,

$$1 < |\Lambda_{min}| \leq |\Lambda(x)| \leq |\Lambda_{max}| \,, \tag{14.4}$$

and the area of each strip in (14.2) is bounded by $|\Lambda_{max}^{-n}| \leq |\mathcal{M}_i| \leq |\Lambda_{min}^{-n}|$. Replacing $|\mathcal{M}_i|$ in (14.2) by its over (under) estimates in terms of $|\Lambda_{max}|$, $|\Lambda_{min}|$ immediately leads to exponential bounds $(2/|\Lambda_{max}|)^n \leq \hat{\Gamma}_n \leq (2/|\Lambda_{min}|)^n$, that is

$$\ln|\Lambda_{max}| \geq -\frac{1}{n}\ln\hat{\Gamma}_n + \ln 2 \geq \ln|\Lambda_{min}|\,. \tag{14.5}$$

The argument based on (14.5) establishes only that the sequence $\gamma_n = -\frac{1}{n}\ln\Gamma_n$ has a lower and an upper bound for any n. In order to prove that γ_n converge to the limit γ, we first show that for hyperbolic systems the sum over survivor intervals (14.2) can be replaced by the sum over periodic orbit stabilities. By (14.4) the size of $\mathcal{M}_i$ strip can be bounded by the stability Λ_i of ith periodic point:

$$\mathcal{C}_1\frac{1}{|\Lambda_i|} < \frac{|\mathcal{M}_i|}{|\mathcal{M}|} < \mathcal{C}_2\frac{1}{|\Lambda_i|}\,, \tag{14.6}$$

for any periodic point i of period n, with constants $\mathcal{C}_j$ 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 ith strip is better and better approximated by by the derivaties evaluated on the periodic point within the strip. Hence the survival probability can be bounded close to the cycle point stability sum

$$\hat{\mathcal{C}}_1\,\Gamma_n < \sum_i^{(n)}\frac{|\mathcal{M}_i|}{|\mathcal{M}|} < \hat{\mathcal{C}}_2\,\Gamma_n \quad , \tag{14.7}$$

where $\Gamma_n = \sum_i^{(n)} 1/|\Lambda_i|$ is the asymptotic trace sum (7.22). In this way we have established that for hyperbolic systems the survival probability sum (14.2) can be replaced by the periodic orbit sum (7.22).

We conclude that for hyperbolic, locally unstable flows the fraction (14.1) of initial x whose trajectories remain trapped within $\mathcal{M}$ up to time t is expected to decay exponentially,

$$\Gamma_{\mathcal{M}}(t) \;\propto\; e^{-\gamma t}\,,$$

where γ is the asymptotic *escape rate* defined by

$$\gamma = -\lim_{t\to\infty}\frac{1}{t}\ln\Gamma_{\mathcal{M}}(t)\,. \tag{14.8}$$

14.1 on p. 331

5.4 on p. 113

Figure 14.1: Johannes Kepler contemplating the bust of Mandelbrot, after Rembrandt's "Aristotle contemplating the bust of Homer" (Metropolitan Museum, New York).
(in order to illustrate the famed New York Times Science section quote!)

14.1.1 Periodic orbit averages

We now refine the reasoning of sect. 14.1. Consider the trace (7.6) in the asymptotic limit (7.21):

$$\mathrm{tr}\,\mathcal{L}^n = \int dx\,\delta(x-f^n(x))\;e^{\beta A^n(x)} \approx \sum_i^{(n)} \frac{e^{\beta A^n(x_i)}}{|\Lambda_i|}\,.$$

The factor $1/|\Lambda_i|$ was interpreted in (14.2) as the area of the ith 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, fig. 1.8, with the volume of the ith 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 ith 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. 10.6.2 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. Equiped 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 $\langle 1\rangle = 1$:

$$\left\langle e^{\beta A^n}\right\rangle_n \approx \frac{\sum_i^{(n)} e^{\beta A^n(x_i)}/|\Lambda_i|}{\sum_i^{(n)} 1/|\Lambda_i|} = \sum_i^{(n)} \mu_i\, e^{\beta A^n(x_i)}\,. \tag{14.9}$$

Here μ_i is the *normalized natural measure*

$$\sum_i^{(n)} \mu_i = 1\,, \qquad \mu_i = e^{n\gamma}/|\Lambda_i|\,, \tag{14.10}$$

correct both for the closed systems as well as the open systems of sect. 6.1.3.

Unlike brute numerical slicing of the integration space into an arbitrary lattice (for a critique, see sect. 9.5), 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 μ_i.

14.1.2 Unstable periodic orbits are dense

(L. Rondoni and P. Cvitanović)

Our goal in sect. 6.1 was to evaluate the space and time averaged expectation value (6.9). An average over all periodic orbits can accomplish the job only if the periodic orbits fully explore the asymptotically accessible phase 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 phase 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, that is one expects a periodic point of period n in cell i. (This is by no means guaranteed to always 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.) 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 (because of the expansiveness of the map) orbits of larger and larger period are densely interspersed in the asymptotic non–wandering set.

14.2 Flow conservation sum rules

If the dynamical system is bounded, all trajectories remain confined for all times, escape rate (14.8) equals $\gamma = -s_0 = 0$, and the leading eigenvalue (**??**) of the Perron-Frobenius operator (5.10) 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:

$$\begin{aligned} 1/\zeta(0,0) &= 1+\sum_{\pi}{}' \frac{(-1)^k}{|\Lambda_{p_1}\cdots\Lambda_{p_k}|} = 0 \\ F(0,0) &= 1-\sum_{n=1}^{\infty} c_n(0,0) = 0 \end{aligned} \tag{14.11}$$

obtained by setting $s = 0$ in (13.12), (13.13) 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 tesselate the phase 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 (13.11) yields

$$\mathrm{tr}\,\mathcal{L}^n = \sum_{i\in \mathrm{Fix} f^n} \frac{1}{|\det\,(\mathbf{1}-\mathbf{J}^n(x_i))\,|} = 1 + e^{s_1 n} + \ldots \ . \tag{14.12}$$

For flows one can apply this rule by grouping together cycles from $t = T$ to $t = T + \Delta T$

$$\begin{aligned} \frac{1}{\Delta T}\sum_{p,r}^{T\le rT_p\le T+\Delta T} \frac{T_p}{\left|\det\,\left(\mathbf{1}-\mathbf{J}_p^r\right)\right|} &= \frac{1}{\Delta T}\int_T^{T+\Delta T} dt\,\left(1+e^{s_1 t}+\ldots\right) \\ &= 1+\frac{1}{\Delta T}\sum_{\alpha=1}^{\infty}\frac{e^{s_\alpha T}}{s_\alpha}\left(e^{s_\alpha \Delta T}-1\right) \approx 1+e^{s_1 T}+\cdots \end{aligned} \tag{14.13}$$

As is usual for the the fixed level trace sums, the convergence of (14.12) is controled by the gap between the leading and the next-to-leading eigenvalues of the evolution operator.

14.3 Correlation functions

The *time correlation function* $C_{AB}(t)$ of two observables A and B along the trajectory $x(t) = f^t(x_0)$ is defined as

$$C_{AB}(t;x_0) \;=\; \lim_{T\to\infty}\frac{1}{T}\int_0^T d\tau A(x(\tau+t))B(x(\tau))\,, \qquad x_0 = x(0)\,. \tag{14.14}$$

If the system is ergodic, with invariant continuous measure $\varrho(x)dx$, then correlation functions do not depend on x_0 (apart from a set of zero measure), and may be computed by a phase average as well

$$C_{AB}(t) \;=\; \int_{\mathcal{M}} dx_0\; \varrho(x_0)A(f^t(x_0))B(x_0)\,. \tag{14.15}$$

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

$$\hat{C}_{AB} := C_{AB} - \langle A\rangle\,\langle B\rangle \tag{14.16}$$

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 $\mathcal{L}$. We can write (14.15) as

$$\tilde{C}_{AB}(t) \;=\; \int_{\mathcal{M}} dx \int_{\mathcal{M}} dy\; A(y)B(x)\varrho(x)\delta(y - f^t(x)),$$

and recover the evolution operator

$$\tilde{C}_{AB}(t) \;=\; \int_{\mathcal{M}} dx \int_{\mathcal{M}} dy\; A(y)\mathcal{L}^t(y,x)B(x)\varrho(x)$$

We also recall that in sect. 5.1 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 $\mathcal{L}$:

$$B(x)\varrho(x) = \sum_{\alpha=0}^{\infty} c_\alpha\varphi_\alpha(x),$$

where $\varphi_0 = \varrho(x)$. 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

$$\tilde{C}_{AB}(t) = \sum_{\alpha \neq 0} e^{-s_\alpha t} c_\alpha \int_{\mathcal{M}} dy\, A(y)\varphi_\alpha(y). \tag{14.17}$$

14.2 on p. 331

We see immediately that if the spectrum has a *gap*, that is the second largest leading eigenvalue is isolated from the largest eigenvalue ($s_0 = 0$) then (14.17) implies an *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 particular observables influences just the prefactor.

The importance of correlation functions, beyond the mentioned theoretical features, is that they are often accessible from time series measurable in laboratory experiments and numerical simulations: moreover they are linked to transport exponents.

14.4 Trace formulas *vs.* level sums

Trace formulas (7.9) and (7.19) diverge precisely where one would like to use them, at s equal to eigenvalues s_α. Instead, one can proceed as follows; according to (7.23) the "level" sums (all symbol strings of length n) are asymptotically going like $e^{s_0 n}$

$$\sum_{i \in \mathrm{Fix} f^n} \frac{e^{\beta A^n(x_i)}}{|\Lambda_i|} = e^{s_0 n}\,,$$

so an nth order estimate $s_{(n)}$ is given by

$$1 = \sum_{i \in \mathrm{Fix} f^n} \frac{e^{\beta A^n(x_i)} e^{-s_{(n)} n}}{|\Lambda_i|} \tag{14.18}$$

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 preasymptotic sequence of finite estimates $s_{(n)}$ is not unique, because the sums Γ_n depend on how we define the escape region, and because in general the areas $\mathcal{M}_i$ in the sum (14.2) 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 (14.8): $s_{(n)} \to s_{(n)} - \ln\alpha/n$. This can be partially fixed by defining a level average

$$\left\langle e^{\beta A(s)} \right\rangle_{(n)} := \sum_{i\in \mathrm{Fix} f^n} \frac{e^{\beta A^n(x_i)} e^{sn}}{|\Lambda_i|} \tag{14.19}$$

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 (14.18), 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 under *all* smooth nonlinear conjugacies $x \to h(x)$, not only linear rescalings, and require no $n \to \infty$ extrapolations. Comparing with the cycle expansions (13.5) 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 $\left|\det\left(\mathbf{1}-\mathbf{J}_p\right)\right| = \left|\det\left(\mathbf{1}-\mathbf{J}^{T_p}(x)\right)\right|$ 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 Jacobian matrix $\mathbf{J}_p$ 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, $\left|\det\left(\mathbf{1}-\mathbf{J}_p\right)\right|$ is invariant under smooth coordinate transformations.

14.4.1 Equipartition measures

There exist many strange sets which cannot be partitioned by the topology of a dynamical flow: some well known examples are the Mandelbrot set, the

period doubling repeller and the probabilistically generated fractal aggregates. In such cases the choice of measure is wide open. One easy choice is the *equipartition* or *cylinder* measure: given a symbolic dynamics partition, weigh all symbol sequences of length n equally. Given a symbolic dynamics, the equipartition measure is easy to implement: the rate of growth of the number of admissible symbol sequences K_n with the sequence length n is given by the topological entropy h (discussed in sect. 11.1) and the equipartition measure for the ith region $\mathcal{M}_i$ is simply

$$\Delta\mu_i = 1/K_n \to e^{-nh} \,. \tag{14.20}$$

The problem with the equipartition measure is twofold: it usually has no physical basis, and it is not an intrinsic invariant property of the strange set, as it depends on the choice of a partition. One is by no means forced to use either the natural or the equipartition measure; there is a variety of other choices, depending on the problem. Also the stability eigenvalues Λ_i need not refer to motion in the dynamical space; in more general settings it can be a renormalization scaling function (sect. **??**), or even a scaling function describing a non–wandering set in the parameter space (sect. 19.3).

Commentary

Remark **14.1** Nonhyperbolic measures. $\mu_i = 1/|\Lambda_i|$ is the natural measure only for the strictly hyperbolic systems. For non-hyperbolic systems, the measure develops folding 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 16), the measure develops a square-root singularity on the $\overline{0}$ cycle:

$$\mu_0 = \frac{1}{|\Lambda_0|^{1/2}} \,. \tag{14.21}$$

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, and we refer the reader to the literature [19, 15, 12, 23].

Remark **14.2** 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 (14.13) seems to have be discussed for the first time by Hannay and Ozorio de Almeida [1, 26]. 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. [3], is that their evaluation *does not* require any explicit construction of the (coordinate dependent) eigenfunctions of the Perron-Frobenius operator (that is, the natural measure ρ_0).

Remark **14.3** The choice of observables We have been quite sloppy on the mathematical side, as in discussing the spectral features of $\mathcal{L}$ the choice of the function space is crucial (especially when one is looking beyond the dominant eigenvalue). As a matter of fact in the function space where usually ergodic properties are defined, $L^2(d\mu)$ there is no gap, due to unitarity property of the Koopman operator: this means that there exist (ugly yet summable) functions for which no exponential decay is present even if the Fredholm determinant has isolated zeroes. A particularly nice example is worked out in [22], and a more mathematical argument is presented in [23].

Remark **14.4** Lattice models The relationship between the spectral gap and exponential decay properties is very well known in the statistical mechanical framework, where one deals with spatial correlations in lattice systems and links them to the gap of the transfer matrix.

Remark **14.5** Role of noise in dynamical systems. In most practical applications in addition to the chaotic deterministic dynamics there is always an additional external noise. The noise can be characterized by its strength σ and distribution. Lyapunov exponents, correlation decay and dynamo rate can be defined in this case the same way as in the deterministic case. We can think that noise completely destroys the results derived here. However, one can show that the deterministic formulas remain valid until the noise level is small. A small level of noise even helps as it makes the dynamics ergodic. Non-communicating parts of the phase space become weakly connected due to the noise. This is a good argument to explain why periodic orbit theory works in non-ergodic systems. For small amplitude noise one can make a noise expansion

$$\overline{\lambda} = \overline{\lambda}_0 + \overline{\lambda}_1\sigma^2 + \overline{\lambda_2}\sigma^4 + ...,$$

around the deterministic averages λ_0. The expansion coefficients $\overline{\lambda_1}, \overline{\lambda_2}, ...$ can also be expressed via periodic orbit formulas. The calculation of these coefficients is one of the challenges facing periodic orbit theory today.

Résumé

We conclude this chapter by a general comment on the relation of the finite trace sums such as (14.2) to the spectral determinants and dynamical zeta functions. One might be tempted to believe that given a deterministic rule, a sum like (14.2) could be evaluated to any desired precision. For short finite times this is

indeed true: every region $\mathcal{M}_i$ in (14.2) 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 (14.2) 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 preasymptotic sequence of finite estimates γ_n is not unique, because the sums Γ_n depend on how we define the escape region, and because in general the areas $\mathcal{M}_i$ in the sum (14.2) should be weighted by the density of initial x_0.

In contrast, the dynamical zeta functions and spectral determinants are already invariant under *all* smooth nonlinear conjugacies $x \to h(x)$, not only linear rescalings, and require no $n \to \infty$ extrapolations.

References

[14.1] F. Christiansen, G. Paladin and H.H. Rugh, *Phys. Rev. Lett.* **65**, 2087 (1990).

Exercises

14.1 Escape rate of the logistic map.

(a) Calculate the fraction of trajectories remaining trapped in the interval $[0, 1]$ for the logistic map

$$f(x) = a(1 - 4(x - 0.5)^2), \tag{14.22}$$

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 $a_c = 1$?

14.2 Four scale map decay. Compute the second largest eigenvalue of the Perron-Frobenius operator for the four scale map

$$f(x) = \begin{cases} a_1 x & \text{if } 0 < x < b/a_1, \\ (1-b)((x - b/a_1)/(b - b/a_1)) + b & \text{if } b/a_1 < x < b, \\ a_2(x - b) & \text{if } b < x < b + b/a_2, \\ (1-b)((x - b - b/a_2)/(1 - b - b/a_2)) + b & \text{if } b + b/a_2 < x < 1. \end{cases} \tag{14.23}$$

14.3 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

$$f(x) = 0.1218x(1-x)(1-0.6x)$$

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 13.2).

(e) for the 2-branch flow conserving map

$$\begin{aligned} f_0(x) &= \frac{h-p+\sqrt{(h-p)^2+4hx}}{2h}\,, \qquad x \in [0,p] \\ f_1(x) &= \frac{h+p-1+\sqrt{(h+p-1)^2+4h(x-p)}}{2h}\,, \qquad x \in [p,1] \end{aligned} \tag{14.24}$$

This is a nonlinear perturbation of ($h = 0$) Bernoulli map (9.10); 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 (6.23) (hint: rescale δx every so often, to avoid numerical overflows),

(g) iterated stability (6.27).

How good is the numerical accuracy compared with the periodic orbit theory predictions?

Chapter 15

Thermodynamic formalism

> So, naturalists observe, a flea hath smaller fleas that on him prey; and those have smaller still to bite 'em; and so proceed ad infinitum.
>
> Jonathan Swift

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 when the concept of fractals and multifractals has been introduced. The formalism made it possible to calculate various fractal dimensions in an elegant way and become a standard instrument in a wide range of scientific fields. Next we sketch the main ideas of this theory and show how periodic orbit theory helps to carry out calculations.

15.1 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 (14.10) the most important is the natural measure, which is the probability that a non-periodic 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 phase space:

$$\sum_{s_1 s_2 \ldots s_n} \mu(s_1 s_2 \ldots s_n) = 1 \tag{15.1}$$

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

$$\lambda(s_1 s_2 \ldots s_n) = -\frac{1}{n} \log \mu(s_1 s_2 \ldots s_n). \tag{15.2}$$

To get full information on the distribution of the natural measure in the symbolic space we can study the distribution of exponents. Let the number of symbol sequences of length n with exponents between λ and $\lambda + d\lambda$ be given by $N_n(\lambda) d\lambda$. For large n the number of such sequences increases exponentially. The rate of this exponential growth can be characterized by $g(\lambda)$ such that

$$N_n(\lambda) \sim \exp(n g(\lambda)).$$

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

$$Z_n(\beta) = \sum_{s_1 s_2 \ldots s_n} \mu^{\beta}(s_1 s_2 \ldots s_n), \tag{15.3}$$

as they are obviously motivated by Gibbs type partition sums of statistical mechanics. The parameter β plays the role of inverse temperature $1/k_B T$ and $E(s_1 s_2 \ldots s_n) = -\log \mu(s 1 s_2 \ldots s_n)$ is the energy associated with the microstate

labelled by $s_1 s_2 ... s_n$ We are tempted also to introduce something analogous with the Free energy. In dynamical systems this is called the Rényi entropy [21] defined by the growth rate of the partition sum

$$K_\beta = \lim_{n\to\infty} \frac{1}{n} \frac{1}{1-\beta} \log \left(\sum_{s_1 s_2 ... s_n} \mu^\beta(s_1 s_2 ... s_n) \right). \tag{15.4}$$

In the special case $\beta \to 1$ we get Kolmogorov's entropy

$$K_1 = \lim_{n\to\infty} \frac{1}{n} \sum_{s_1 s_2 ... s_n} -\mu(s_1 s_2 ... s_n) \log \mu(s_1 s_2 ... s_n),$$

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 (15.2). For large n we can replace $N_n(\lambda)$ with its asymptotic form

$$Z_n(\beta) \sim \int d\lambda \exp(ng(\lambda)) \exp(-n\lambda\beta).$$

For large n this integral is dominated by contributions from those λ^* which maximize the exponent

$$g(\lambda) - \lambda\beta.$$

The exponent is maximal when the derivative of the exponent vanishes

$$g'(\lambda^*) = \beta. \tag{15.5}$$

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 (15.4) we can now connect the Rényi entropies and $g(\lambda)$

$$(\beta - 1)K_\beta = \lambda^*(\beta)\beta - g(\lambda^*(\beta)). \tag{15.6}$$

Equations (15.5) and (15.6) 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

$$g(\lambda) = \lambda\beta^*(\lambda) - (\beta^*(\lambda) - 1)K_{\beta^*(\lambda)}, \tag{15.7}$$

where now $\beta^*(\lambda)$ can be determined from

$$\frac{d}{d\beta^*}[(\beta^* - 1)K_{\beta^*}] = \lambda. \tag{15.8}$$

Obviously, if we can determine the Rényi entropies we can recover the distribution of probabilities from (15.7) and (15.8).

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 (14.10)

$$\mu(s_1, ..., s_n) \approx \frac{e^{n\gamma}}{|\Lambda_{\overline{s_1 s_2 ... s_n}}|}. \tag{15.9}$$

The partition sum (15.3) now reads

$$Z_n(\beta) \approx \sum_i \frac{e^{n\beta\gamma}}{|\Lambda_i|^\beta}, \tag{15.10}$$

where the summation goes for periodic orbits of length n. We can define the characteristic function

$$\Omega(z, \beta) = \exp\left(-\sum_n \frac{z^n}{n} Z_n(\beta)\right). \tag{15.11}$$

According to (15.4) for large n the partition sum behaves as

$$Z_n(\beta) \sim e^{-n(\beta-1)K_\beta}. \tag{15.12}$$

Substituting this into (15.11) 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 (15.10) into (15.11) and introducing primitive 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

$$P(\beta) = (\beta-1)K_\beta + \beta\gamma, \tag{15.13}$$

since the leading zero of the zeta function is the pressure. The Rényi entropies K_β, hence the distribution of the exponents $g(\lambda)$ as well, can be calculated via finding the leading eigenvalue of the operator (G.4).

From (15.13) we can get all the important quantities of the thermodynamic formalism. For $\beta = 0$ we get the topological entropy

$$P(0) = -K_0 = -h_{top}. \tag{15.14}$$

For $\beta = 1$ we get the escape rate

$$P(1) = \gamma. \tag{15.15}$$

Taking the derivative of (15.13) in $\beta = 1$ we get Pesin's formula [2] connecting Kolmogorov's entropy and the Lyapunov exponent

$$P'(1) = \overline{\lambda} = K_1 + \gamma. \tag{15.16}$$

15.1 on p. 343

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 fig. 15.2 we show a typical pressure and $g(\lambda)$ curve computed for the two scale tent map of Exercise 15.4. We have to mention, that all typical hyperbolic dynamical system produces a similar parabola like curve. Although this is somewhat boring we can interpret it like a sign of a high level of universality: The exponents λ have a sharp distribution around the most probable value. The most probable value is $\lambda = P'(0)$ and $g(\lambda) = h_{top}$ is the topological entropy. The average value in closed systems is where $g(\lambda)$ touches the diagonal: $\overline{\lambda} = g(\overline{\lambda})$ and $1 = g'(\overline{\lambda})$.

Next, we are looking at the distribution of trajectories in real space.

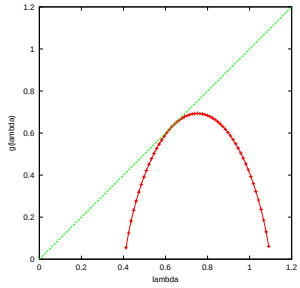


Figure 15.1:

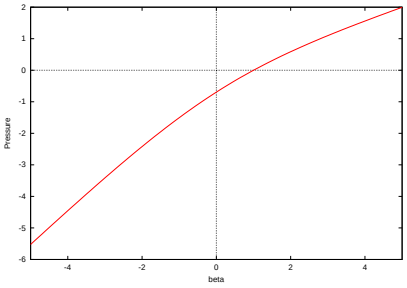


Figure 15.2: $g(\lambda)$ and $P(\beta)$ for the map of Exercise 15.4 at $a = 3$ and $b = 3/2$. See Solutions K for calculation details.

15.2 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 bounce form two strips. Then these strips are subdivided into an infinite hierarchy of substrips as we follow trajectories which do not leave the system after more and more bounces. The finer strips are similar to strips on a larger scale. Objects with such self similar properties are called *fractals*.

We can characterize fractals via their local scaling properties. The first step is to draw a uniform grid on the surface of section. We can look at various measures in the square boxes of the grid. The most interesting measure is again the natural measure located in the box. By decreasing the size of the grid ϵ the measure in a given box will decrease. If the distribution of the measure is smooth then we expect that the measure of the i-th box is proportional with the dimension of the section

$$\mu_i \sim \epsilon^d.$$

If the measure is distributed on a hairy object like the repeller we can observe unusual scaling behavior of type

$$\mu_i \sim \epsilon^{\alpha_i},$$

where α_i is the local "dimension" or Hölder exponent of the the object. As α is not necessarily an integer here we are dealing with objects with fractional dimensions. We can study the distribution of the measure on the surface of section by looking

at the distribution of these local exponents. We can define

$$\alpha_i = \frac{\log \mu_i}{\log \epsilon},$$

the local Hölder exponent and then we can count how many of them are between α and $\alpha + d\alpha$. This is $N_\epsilon(\alpha)d\alpha$. Again, in smooth objects this function scales simply with the dimension of the system

$$N_\epsilon(\alpha) \sim \epsilon^{-d},$$

while for hairy objects we expect an α dependent scaling exponent

$$N_\epsilon(\alpha) \sim \epsilon^{-f(\alpha)}.$$

$f(\alpha)$ can be interpreted [8] as the dimension of the points on the surface of section with scaling exponent α. We can calculate $f(\alpha)$ with the help of partition sums as we did for $g(\lambda)$ in the previous section. First we define

$$Z_\epsilon(q) = \sum_i \mu_i^q. \tag{15.17}$$

Then we would like to determine the asymptotic behavior of the partition sum characterized by the $\tau(q)$ exponent

$$Z_\epsilon(q) \sim \epsilon^{-\tau(q)}.$$

The partition sum can be written in terms of the distribution function of α-s

$$Z_\epsilon(q) = \int d\alpha N_\epsilon(\alpha)\epsilon^{q\alpha}.$$

Using the asymptotic form of the distribution we get

$$Z_\epsilon(q) \sim \int d\alpha \epsilon^{q\alpha - f(\alpha)}.$$

As ϵ goes to zero the integral is dominated by the term maximizing the exponent. This α^* can be determined from the equation

$$\frac{d}{d\alpha^*}(q\alpha^* - f(\alpha^*)) = 0,$$

leading to

$$q = f'(\alpha^*).$$

Finally we can read off the scaling exponent of the partition sum

$$\tau(q) = \alpha^* q - f(\alpha^*).$$

In a uniform fractal characterized by a single dimension both α and $f(\alpha)$ collapse to $\alpha = f(\alpha) = D$. The scaling exponent then has the form $\tau(q) = (q-1)D$. In case of non uniform fractals we can introduce generalized dimensions [10] D_q via the definition

$$D_q = \tau(q)/(q-1).$$

Some of these dimensions have special names. For $q = 0$ the partition sum (15.17) 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 [17] of the distribution, hence its name is *information dimension*.

Using equisize grids is impractical in most of the applications. Instead, we can rewrite (15.17) into the more convenient form

$$\sum_i \frac{\mu_i^q}{\epsilon^\tau(q)} \sim 1. \tag{15.18}$$

If we cover the ith branch of the fractal with a grid of size l_i instead of ϵ we can use the relation [9]

$$\sum_i \frac{\mu_i^q}{{l_i}^\tau(q)} \sim 1, \tag{15.19}$$

the non-uniform grid generalization of 15.18. 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 situating in them

$$l_i \sim \frac{1}{|\Lambda_i|}.$$

Then using this relation and the periodic orbit expression of the natural measure we can write (15.19) into the form

$$\sum_i \frac{e^{q\gamma n}}{|\Lambda_i|^{q-\tau(q)}} \sim 1, \tag{15.20}$$

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 (15.20) 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

$$P(q-(q-1)D_q) = q\gamma. \tag{15.21}$$

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 (15.21) 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 (15.21) 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

$$D_1 = 1 - \gamma/\overline{\lambda}. \tag{15.22}$$

If the system is bound ($\gamma = 0$) the information dimension and all other dimensions are $D_q = 1$. Also since D_10 is positive (15.22) proves that the Lyapunov exponent must be larger than the escape rate $\overline{\lambda} > \gamma$ in general.

15.4 on p. 344
15.5 on p. 344
15.6 on p. 345

Commentary

Remark **15.1** Mild phase transition In non-hyperbolic systems the formulas derived in this chapter should be modified. As we mentioned in 14.1 in non-hyperbolic systems the periodic orbit expression of the measure can be

$$\mu_0 = e^{\gamma n}/|\Lambda_0|^{\delta},$$

where δ can differ from 1. Usually it is 1/2. For sufficiently *negative* β the corresponding term $1/|\Lambda_0|^{\beta}$ can dominate (15.10) while in (15.3) $e^{\gamma n}/|\Lambda_0|^{\delta\beta}$ plays no dominant role. In this case the pressure as a function of β can have

a kink at the critical point $\beta = \beta_c$ where $\beta_c \log |\Lambda_0| = (\beta_c - 1)K_{\beta_c} + \beta_c \gamma$. For $\beta < \beta_c$ the pressure and the Rényi entropies differ

$$P(\beta) \neq (\beta - 1)K_\beta + \beta\gamma.$$

This phenomena is called phase transition. This is however not a very deep problem. We can fix the relation between pressure and the entropies by replacing $1/|\Lambda_0|$ with $1/|\Lambda_0|^\delta$ in (15.10).

Remark **15.2** 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 (15.10) and (15.3). 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.

Remark **15.3** Multifractals. For reasons that remain mysterious to the authors - perhaps so that Mandelbrot can refer to himself both as the mother of fractals and the grandmother of multifractals - some physics literature referes to any fractal generated by more than one scale as a "multi"-fractal. This usage seems to divide fractals into 2 classes; one consisting essentially of the above Cantor set and the Serapinski gasket, and the second consisting of anything else, including all cases of physical interest.

Résumé

In this chapter we have shown that thermodynamic quantities and various fractal dimensions can be expressed in terms of the pressure function. The pressure function is the leading eigenvalue of the operator which generates the Lyapunov exponent. In the Lyapunov case β is just an auxiliary variable. In thermodynamics it plays an essential role. The good news of the chapter is that the distribution of locally fluctuating exponents should not be computed via making statistics. We can use cyclist formulas for determining the pressure. Then the pressure can be found using short cycles + curvatures. Here the head reach 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.

Exercises

15.1 Thermodynamics in higher dimensions Introduce the time averages of the eigenvalues of the Jacobian

$$\lambda_i = \lim_{t\to\infty} \frac{1}{t} \log |\Lambda_i^t(x_0)|, \tag{15.23}$$

as a generalization of (6.27).

Show that in higher dimensions Pesin's formula is

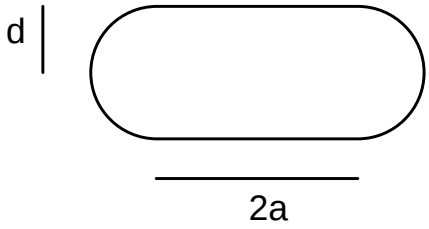
$$K_1 = \sum_i \lambda_i - \gamma, \tag{15.24}$$

where the summation goes for the positive λ_i-s only. (Hint: Use the higher dimensional generalization of (14.10)

$$\mu_i = e^{n\gamma} / |\prod_j \Lambda_{i,j}|,$$

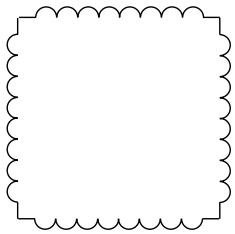
where the product goes for the expanding eigenvalues of the Jacobian of the periodic orbit.

15.2 Bunimovich stadium Kolmogorov entropy. Take for definitiveness $a = 1.6$ and $d = 1$ in the Bunimovich stadium of exercise 4.3,



estimate the Lyapunov exponent by averaging over a very long trajectory. Biham and Kvale [**?**] 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) $h \approx 1.15 \pm .03$.

15.3 Entropy of rugged-edge billiards. Take a semi-circle of diameter ε and replace the sides of a unit square by $\lfloor 1/\varepsilon \rfloor$ catenated copies of the semi-circle.



(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 Bunimovich stadium are far apart, say L, the entropy for the flow decays as

$$K_1 \to \frac{2 \ln L}{\pi L}\,.$$

15.4 Two scale map Compute all those quantities - dimensions, escape rate, entropies, etc. - for the repeller of the one dimensional map

$$f(x) = \begin{cases} 1 + ax & \text{if} \quad x < 0, \\ 1 - bx & \text{if} \quad x > 0. \end{cases} \tag{15.25}$$

where a and b are larger than 2. Compute the fractal dimension, plot the pressure and compute the $f(\alpha)$ spectrum of singularities. Observe how K_1 may be obtained directly from (**??**).

15.5 Four scale map Compute the Rényi entropies and $g(\lambda)$ for the four scale map

$$f(x) = \begin{cases} a_1 x & \text{if} \quad 0 < x < b/a_1, \\ (1-b)((x - b/a_1)/(b - b/a_1)) + b & \text{if} \quad b/a_1 < x < b, \\ a_2(x-b) & \text{if} \quad b < x < b + b/a_2, \\ (1-b)((x - b - b/a_2)/(1 - b - b/a_2)) + b & \text{if} \quad b + b/a_2 < x < 1. \end{cases} \tag{15.26}$$

Hint: Calculate the pressure function and use (15.13).

15.6 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.

Chapter 16

Intermittency

Sometimes They Come Back
Stephen King

(R. Artuso, P. Dahlqvist and G. Tanner)

In the theory of chaotic dynamics developed so far we assumed that the evolution operator has a discrete spectra $\{z_0, z_1, z_2, \ldots\}$ given by the zeros of

$$1/\zeta(z) = (\cdots)\prod_k (1 - z/z_k)\,,$$

Such an assumption was based on the tacit premise that the dynamics is everywhere exponentially unstable. Real life is nothing like that - phase spaces are generically infinitely interwoven patterns of stable and unstable behaviors. While the stable ("integrable") and the unstable ("chaotic") behaviors are by now pretty much under control, the borderline marginally stable orbits present many difficult 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.

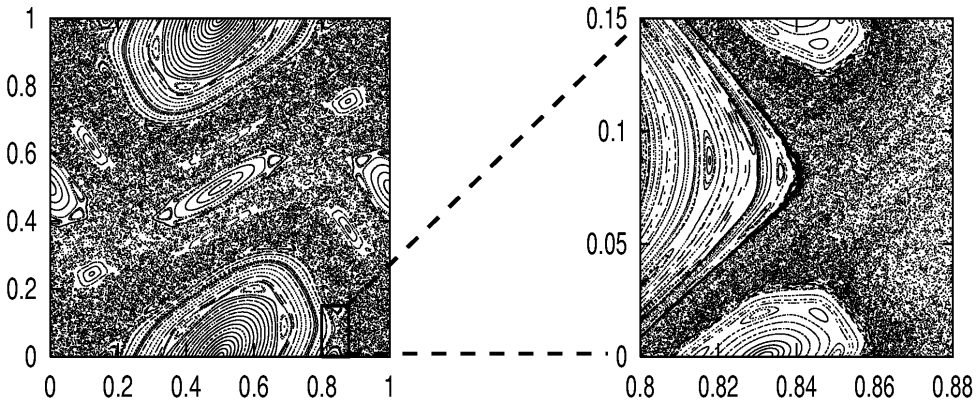


Figure 16.1: Typical phase space for an area-preserving map with mixed phase space dynamics; (here the standard map for k=1.2).

16.1 Intermittency everywhere

With a change in an external parameter, one observes in many fluid dynamics experiments a transition from a regular behavior to a behavior where long time intervals of regular behavior ("laminar phases") are interupted 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 phase space regions. Complete hyperbolicity assumed in (7.5) 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 phase spaces with islands of stability coexisting with hyperbolic regions, see fig. 16.1. 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 interupted 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, as 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 is illustrated by fig. 16.1.

Intermittency is due to the existence of fixed points and cycles of marginal stability (4.59), or (in studies of the onset of intermittency) to the proximity of a nearly marginal complex or unstable orbit. In Hamiltonian systems intermit-

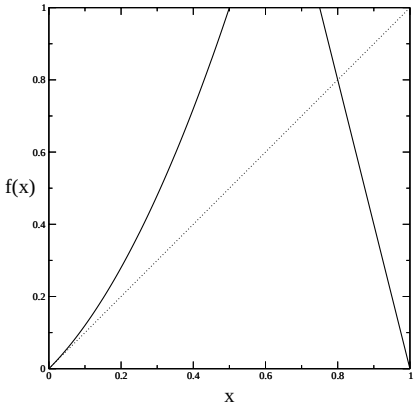


Figure 16.2: A complete binary repeller with a marginal fixed point.

tency 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 fig. 10.11). Following the stretching and folding of the invariant manifolds in time one will inevitably find phase 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 phase space or, in other words, marginal stability. Under small parameter perturbations such neighborhoods undergo tangent birfucations - a stable/unstable pair of periodic orbits is destroyed or created by coalescing into a marginal orbit, so pruning which we encountered first in chapter **??**, and intermittency are two sides of the same coin.
.

☞ sect. 10.7

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 of the form

$$x \mapsto f(x) = x + O(x^{1+s})\,. \tag{16.1}$$

which are expanding almost everywhere except for a single marginally stable fixed point at x=0. Such a map may allow escape, like the map shown in fig. 16.2 or may be bounded like the Farey map (13.26)

$$x \mapsto f(x) = \left\{ \begin{array}{ll} x/(1-x) & x \in [0, 1/2[\\ (1-x)/x & x \in [1/2, 1] \end{array} \right\}\,.$$

introduced in sect. 13.4. Fig. 16.3 compares a trajectory of the (uniformly hyperbolic) tent map (10.15) side by side with a trajectory of the (non-hyperbolic) Farey map. In a stark contrast to the uniformly chaotic trajectory of the tent

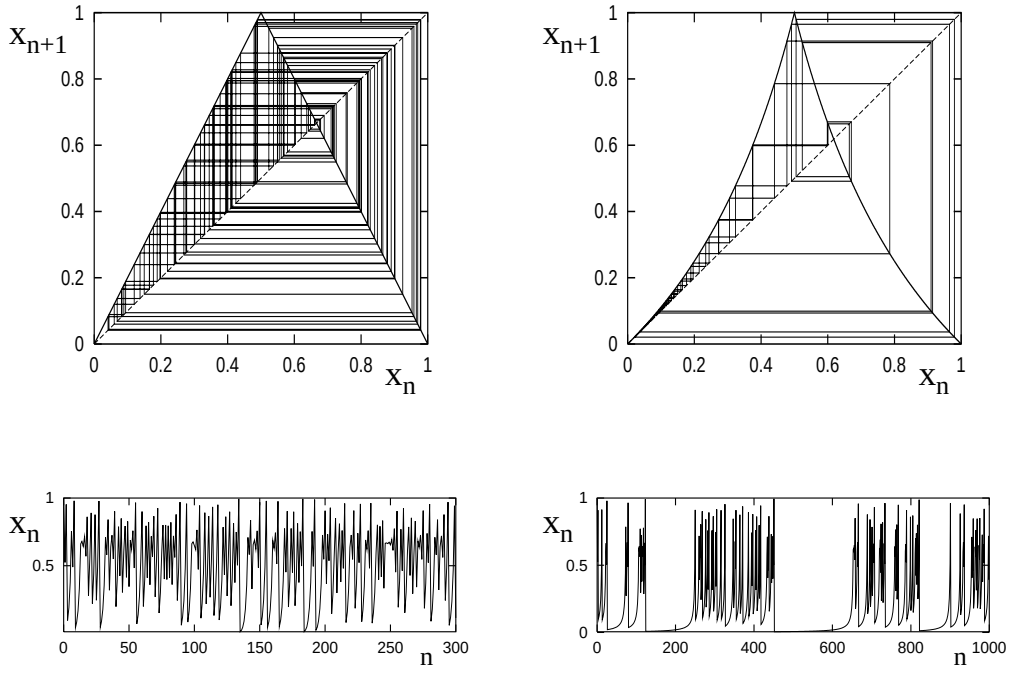


Figure 16.3: (a) A tent map trajectory. (b) A Farey map trajectory.

map, the Farey map trajectory alternates intermittently between slow regular motion of varying length glued to the marginally stable fixed point, and chaotic bursts.

☞ sect. 13.4.3

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 fig. 16.2 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 need to answer are: how does marginal stability affect zeta functions or spectral determinants? And, can we deduce power law decays of correlations from cycle expansions?

In sect. 9.2.2 we saw that marginal stability violates one of the conditions which ensure that the spectral determinant is an entire function. Already the simple fact that the cycle weight $1/|1-\Lambda_p^r|$ in the trace (7.3) or the spectral determinant (8.3) diverges for marginal orbits with $|\Lambda_p| = 1$ tells us that we have to treat these orbits with care. We saw in sect. 13.4 that a cycle expansion for the Farey map based on the binary symbolic dynamics does not reflect the nonuniform distribution of cycle weights of the map; in that example a stability ordered expansion leads to improved convergence properties.

In the following we will take a more systematic approach to incorporate marginal stability into a cycle-expansion. To get to know the difficulties lying ahead, we will first start with a map, which is piecewise linear, but still follows the asymptotics (16.1) in sect. 16.2. We will construct a dynamical zeta function in the usual way without worrying too much about its justification at that stage and show that it has a branch point singularity. We will calculate the rate of escape from our piecewise linear map and find a power law behavior. The worrying comes next: that is, we will argue that dynamical zeta functions in the presence of marginal stability can still be written in terms of periodic orbits exactly in the way as derived in chapters 6 and 14 with one exception: we actually have to exclude the marginal stable fixed point explicitely. This innocent looking step has far reaching consequences; it forces us to change from finite symbolic dynamics to an infinite letter symbol code and demands a reorganisation of the order of summation in the cycle expansion. We will come to these more conceptual issues in sect. 16.2.3

Branch points are typical also for smooth intermittent maps with isolated marginally stable fixed points and cycles. In sect. 16.3, we discuss the cycle expansions and curvature combinations for zeta functions of smooth maps taylored for intermittency. The knowledge of the type of singularity one encounters enables us to construct an efficient resummation method which is presented in sect. 16.3.1.

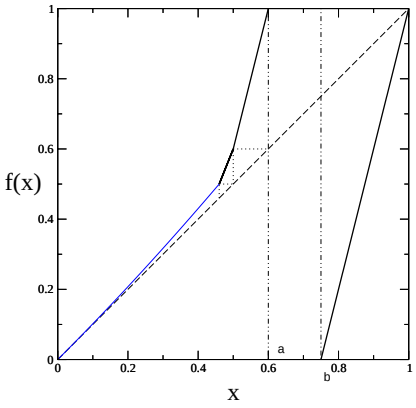


Figure 16.4: A piecewise linear intermittent map, see (16.2).

Finally, in sect. 16.4, we discuss a probabilistic method that yields approximate dynamical zeta functions and provides valuable information about more complicated systems, such as billiards.

16.2 Intermittency for beginners

Intermittency does not only present us with a large repertoire of interesting dynamics, it is also at the root of problems, such as slow convergence of cycle expansions or pruning. 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 a carefully constructed piecewise linear model first. From there we will move on to the more general case of a smooth intermittend map which will be discussed in sect. 16.3.

16.2.1 A toy map

The binary shift map is an idealised example of a hyperbolic map. To study intermittency we will now construct a piecewise linear model, which can be thought of as 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

$$f(x) = \begin{cases} f_0(x) & x \in \mathcal{M}_0 = [0, a] \\ f_1(x) & x \in \mathcal{M}_1 =]b, 1] \end{cases} . \tag{16.2}$$

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 fig. 16.4).

We will choose the right branch to be expanding and linear, that is,

$$f_1(x) = \frac{x-b}{1-b}\,.$$

Next, we will construct the left branch in a way, which will allow us to model the intermittent behaviour (16.1) 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

$$\mathcal{M}_n =]q_n, q_{n-1}] \qquad \text{and} \qquad \mathcal{M}_0 = \bigcup_{n=2}^{\infty} \mathcal{M}_n. \tag{16.3}$$

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 $\mathcal{M}_n = (f_0^{-1})^{n-1}([a,1])$.

This fixes the map for any given sequence $\{q_n\}$. The last condition ensures the existence of a simple Markov partition. The slopes of the various linear segments are

$$\begin{aligned} f_0'(x) &= \frac{f_0(q_{n-1}) - f_0(q_n)}{q_{n-1} - q_n} = \frac{|\mathcal{M}_{n-1}|}{|\mathcal{M}_n|} \quad \text{for} \quad x \in \mathcal{M}_n \quad \text{and} \quad n \geq 3 \\ f_0'(x) &= \frac{f_0(q_1) - f_0(q_2)}{q_1 - q_2} = \frac{1-a}{|\mathcal{M}_2|} \quad \text{for} \quad x \in \mathcal{M}_2 \\ f_0'(x) &= \frac{1}{1-b} \quad \text{for} \quad x \in \mathcal{M}_1 \end{aligned} \tag{16.4}$$

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 (16.1). An orbit 10^n enters the intervals $\mathcal{M}_1\mathcal{M}_{n+1}, \mathcal{M}_n, \ldots \mathcal{M}_2$ 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

$$\Lambda_{10^n} = f_0'(x_{n+1}) f_0'(x_n) \ldots f_0'(x_2) f_1'(x_1) = \frac{1}{|\mathcal{M}_{n+1}|} \frac{1-a}{1-b} \tag{16.5}$$

with $x_i \in \mathcal{M}_i$. The properties of the map (16.2) are completely determined by the sequence $\{q_n\}$. By choosing $q_n = 2^{-n}$, for example, we recover the uniformly hyperbolic binary shift map. An intermittent map of the form (16.3) having the asymptotic behaviour (16.1) can be constructed by chosing an algebraically decaying sequence $\{q_n\}$ behaving asymptotically like

$$q_n \sim \frac{1}{n^{1/s}}, \tag{16.6}$$

where s is the intermittency exponent in (16.1). Such a partition leads to intervals whose length decreases asymptotically like a power-law, that is,

$$|\mathcal{M}_n| \sim \frac{1}{n^{1+1/s}}\,. \tag{16.7}$$

The stability of periodic orbit families approaching the marginal fixed point, as for example the family of orbits with symbol code 10^n increases in turn only algebraically with the cycle length as can be seen from refeq (16.5).

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 (16.7) 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, see remark 16.3. We will, however, not go along this route here; instead, we will choose a maybe less obvious partition which will simplify the algebra considerably later without loosing any of the key features typical for intermittent systems. We fix the intermittent toy map by specifying the intervals $\mathcal{M}_n$ in terms of gamma functions according to

$$|\mathcal{M}_n| = \mathcal{C}\frac{\Gamma(n+m-1/s-1)}{\Gamma(n+m)} \qquad \text{for} \quad n \geq 2, \tag{16.8}$$

where $m = [1/s]$ denotes the integer part of $1/s$ and $\mathcal{C}$ is a normalization constant fixed by the condition $\sum_{n=2}^{\infty} |\mathcal{M}_n| = q_1 = a$, that is,

$$\mathcal{C} = a\left[\sum_{n=2}^{\infty} \frac{\Gamma(n-1/s)}{\Gamma(n+1)}\right]^{-1}. \tag{16.9}$$

Using Stirling's formula for the Gamma function

$$\Gamma(z) \sim e^{-z} z^{z-1/2}\sqrt{2\pi}(1 + \frac{1}{12z} + \ldots),$$

we find that the intervals decay asymptotically like $n^{-(1+1/s)}$ as required by the condition (16.7).

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. 13.4, that such a cycle expansion converges extremely slow in the presence of marginal stability. The shadowing mechanism between orbits and pseudo-orbits is very inefficient for orbits of the form 10^n with stabilities given by (16.5) due to the marginal stability of the fixed point $\overline{0}$. It is therefore advantagous to choose as the fundamental cycles the family of orbits with code 10^n or equivalently switching 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 $\mathcal{M}_{n-1}$, we get the transformation from a finite alphabet to an infinite alphabet here for free. All periodic orbits entering the left branch at least twice are cancelled exactly by composite orbits and the cycle expanded dynamical zeta function has the simple form

$$\begin{aligned} 1/\zeta(z) &= \prod_{p\neq 0}\left(1 - \frac{z^{n_p}}{|\Lambda_p|}\right) = 1 - \sum_{n=1}^{\infty} \frac{z^n}{|\Lambda_{10^{n-1}}|} \\ &= 1 - (1-b)z - \mathcal{C}\frac{1-b}{1-a}\sum_{n=2}^{\infty}\frac{\Gamma(n+m-1/s-1)}{\Gamma(n+m)}z^n\,. \end{aligned} \tag{16.10}$$

The fundamental term consists here of an infinite sum over algebraically decaying cycle weights. The sum is divergent for $|z| \geq 1$, that is, the cycle expansion does not provide an analytic continuation, here, despite the fact that all curvature terms cancel exactly. We will see that this behavior is due to a branchcut of $1/\zeta$ starting at $z = 1$. We are thus faced with the extra effort to find analytic continuations of sums over algebraically decreasing terms as they appear in (16.10). 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. 16.2.3.

16.2.2 Branch cuts and the escape rate

Starting from the dynamical zeta function (16.10), 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 (16.8). The sum over ratios of Gamma functions in (16.10) can be evaluated analytically by using the following identities valid for $1/s = \alpha > 0$:

- α non-integer

$$(1-z)^\alpha = \sum_{n=0}^{\infty} \frac{\Gamma(n-\alpha)}{\Gamma(-\alpha)\Gamma(n+1)} z^n \tag{16.11}$$

- α integer

$$\begin{aligned} (1-z)^\alpha \log(1-z) &= \sum_{n=1}^{\alpha} (-1)^n c_n z^n \\ &+ (-1)^{\alpha+1} \alpha! \sum_{n=\alpha+1}^{\infty} \frac{(n-\alpha-1)!}{n!} z^n \end{aligned} \tag{16.12}$$

  with

$$c_n = \left(\begin{array}{c} \alpha \\ n \end{array} \right) \sum_{k=0}^{n-1} \frac{1}{\alpha - k} .$$

In order to simplify the notation, we will restrict ourselves for a while to intermittency parameters in the range $1 \leq 1/s < 2$, that is, we have $[1/s] = m = 1$. All what follows can easily be generalized to arbitrary $s > 0$ using equations (16.11) and (16.12). The infinite sum in (16.10) can now be evaluated with the help of (16.11) or (16.12), that is,

$$\sum_{n=2}^{\infty} \frac{\Gamma(n-1/s)}{\Gamma(n+1)} z^n = \left\{ \begin{array}{ll} \Gamma(-\frac{1}{s}) \left[(1-z)^{1/s} - 1 + \frac{1}{s} z \right] & \text{for} \quad 1 < 1/s < 2; \\ (1-z)\log(1-z) + z & \text{for} \quad s = 1 \, . \end{array} \right.$$

The normalization constant $\mathcal{C}$ in (16.8) can be evaluated explicitely using Eq. (16.9) and the dynamical zeta function can be given in closed form. We obtain for $1 < 1/s < 2$

$$1/\zeta(z) = 1 - (1-b)z + \frac{a}{1-1/s} \frac{1-b}{1-a} \left((1-z)^{1/s} - 1 + \frac{1}{s} z \right) . \tag{16.13}$$

and for $s = 1$,

$$1/\zeta(z) = 1 - (1-b)z + a\frac{1-b}{1-a}\left((1-z)\log(1-z) + z\right). \tag{16.14}$$

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 explicitely, 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 approching 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

$$1/\zeta(z) = 1 - (1-b)z + \frac{a}{g_s(1)}\frac{1-b}{1-a}\frac{1}{z^{m-1}}\left((1-z)^{1/s} - g_s(z)\right) \tag{16.15}$$

for non-integer s with $m = [1/s]$ and

$$1/\zeta(z) = 1-(1-b)z+\frac{a}{g_m(1)}\frac{1-b}{1-a}\frac{1}{z^{m-1}}\left((1-z)^m\log(1-z) - g_m(z)\right) \tag{16.16}$$

for $1/s = m$ integer and $g_s(z)$ are polynomials of order $m = [1/s]$ which can be deduced from (16.11) or (16.12). 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. 16.3 that branch cuts of that form are generic for 1-dimensional intermittent maps.

Branch cuts are the all important new feature which is introduced 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? Let's take 'a learning by doing'-approach and calculate the escape from our toy map for $a < b$.

A useful starting point for the calculation of the fraction of survivors after n steps, Γ_n, is the integral representation (8.16), that is

$$\Gamma_n = \frac{1}{2\pi i}\oint_{\gamma_r^-} z^{-n}\left(\frac{d}{dz}\log\zeta^{-1}(z)\right)dz \tag{16.17}$$

where the contour encircles the origin in negative, that is, in clockwise direction. If the contour is small enough, e.g. lies inside the unit cicle $|z| = 1$, we may write the logarithmic derivative of $\zeta^{-1}(z)$ as a convergent sum over all periodic orbits. Integrals and sums can be interchanged, the integrals can be solved term by

Figure 16.5: The survival probability Γ_n can be split into contributions from poles (x) and zeros (o) between the small and the large circle and a contribution from the large circle.

term, and the formula (7.22) is recovered. For hyperbolic maps, cycle expansion methods or other techniques may provide an analytic extension of the dynamical zeta function beyond the leading zero; we may therefore deform the orignal contour into a larger circle with radius R which encircles both poles and zeros of $\zeta^{-1}(z)$, see fig. 16.5. Residue calculus turns this into a sum over the zeros z_α and poles z_β of the dynamical zeta function, that is

$$\Gamma_n = \sum_{|z_\alpha|<R}^{\text{zeros}} \frac{1}{z_\alpha^n} - \sum_{|z_\beta|<R}^{\text{poles}} \frac{1}{z_\beta^n} + \frac{1}{2\pi i} \oint_{\gamma_R^-} dz\, z^{-n} \frac{d}{dz} \log \zeta^{-1}, \tag{16.18}$$

where the last term gives a contribution from a large circle γ_R^-. We thus find exponential decay of Γ_n dominated by the leading zero or pole of $\zeta^{-1}(z)$, see chapter 15.1 for more details.

Things change considerably in the intermittent case. The point $z = 1$ is a branch point singularity and there exists no Taylor series expansion of ζ^{-1} around $z = 1$. Secondly, the path deformation that led us to (16.18) requires more care, as it may 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 point in anti-clockwise direction, see fig. 16.6. We will denote the detour around the cut as γ_{cut}. We may write symbolically

$$\oint_{\gamma_r} = \sum^{\text{zeros}} - \sum^{\text{poles}} + \oint_{\gamma_R} + \oint_{\gamma_{cut}}$$

where the sums include only the zeros and the poles in the area enclosed by the contours. The asymptotics is controlled by the zero, pole or cut, which is closest

Figure 16.6: In the intermittent case the large circle γ_R^- in fig. 16.5 must not cross the branch cut, it has to make the detour γ_{cut}

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 (16.17) at the branch point $z=1$. We restrict ourselves again to the case $1<1/s<2$ first and write the dynamical zeta function (16.13) 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

$$\frac{1}{2\pi i}\oint_{\gamma_{cut}} (1-u)^{-n}\frac{d}{du}\log G(u)du \tag{16.19}$$

where γ_{cut} goes around the cut (that is, 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

$$\frac{d}{du}\log G(u) = \frac{a_1}{a_0} + \frac{1}{s}\frac{b_0}{a_0}u^{1/s-1} + O(u)\,. \tag{16.20}$$

The integrals along the cut may be evaluated using the general formula

$$\frac{1}{2\pi i}\oint_{\gamma_{cut}} u^{\alpha}(1-u)^{-n}du = \frac{\Gamma(n-\alpha-1)}{\Gamma(n)\Gamma(-\alpha)} \sim \frac{1}{n^{\alpha+1}}(1+O(1/n)) \tag{16.21}$$

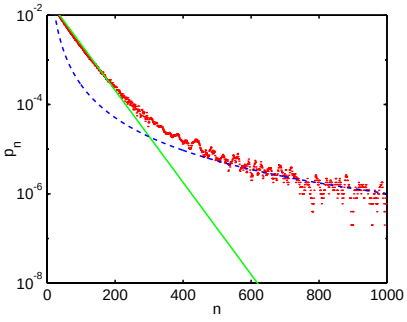


Figure 16.7: The asymptotic escape from an intermittent repeller is a power law. Normally it is preceded by an exponential, which can be related to zeros close to the cut but beyond the branch point $z = 1$, as in fig. 16.6.

which can be obtained by deforming the contour back to a loop around the point $u = 1$, now in positive (anti-clockwise) direction. The contour integral then picks up the $n-1$st term in the Taylor expansion of the function u^α at $u = 1$, cf. (16.11). For the continuous time case the corresponding formula is

$$\frac{1}{2\pi i}\oint_{\gamma_{cut}} z^\alpha e^{zt} dz = \frac{1}{\Gamma(-\alpha)}\frac{1}{t^{\alpha+1}}. \tag{16.22}$$

Plugging (16.20) into (16.19) and using (16.21) we get the asymptotic result

$$\Gamma_n \sim \frac{b_0}{a_0}\frac{1}{s}\frac{1}{\Gamma(1-1/s)}\frac{1}{n^{1/s}} = \frac{a}{s-1}\frac{1-b}{b-a}\frac{1}{\Gamma(1-1/s)}\frac{1}{n^{1/s}}\,. \tag{16.23}$$

We see that, asymptotically, the escape from an intermittent repeller is described by power law decay rather than the exponential decay we are familiar with for hyperbolic maps; a numerical simulation of the power-law escape from an intermittent repeller is shown in fig. 16.7.

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 (16.15). Expanding $\frac{d}{du}\log G(u)$ around $u = 0$, one again obtains leading order contributions according to Eq. (16.20) and the general result follows immediatly using (16.21), that is,

$$\Gamma_n \sim \frac{a}{sg_s(1)}\frac{1-b}{b-a}\frac{1}{\Gamma(1-1/s)}\frac{1}{n^{1/s}}\,. \tag{16.24}$$

Applying the same arguments for integer intermittency exponents $1/s = m$, one obtains

$$\Gamma_n \sim \ (-1)^{m+1} \frac{a}{s g_m(1)} \frac{1-b}{b-a} \frac{m!}{n^m} \,. \tag{16.25}$$

So far, we have considered the survival probability for a repeller, that is we assumed $a < b$. The formulas (16.24) and (16.25) do obviously not apply for the case $a = b$, that is, for the bounded map. The coefficients $a_0 = (b-a)/(1-a)$ in the series representation of $G(u)$ is zero and the expansion of the logarithmic derivative of $G(u)$, Eq. (16.20) is now longer valid. We now get instead

$$\frac{d}{du} \log G(u) = \begin{cases} \frac{1}{u}\left(1 + O(u^{1/s-1})\right) & s < 1 \\ \frac{1}{u}\left(\frac{1}{s} + O(u^{1-1/s})\right) & s > 1 \end{cases} ,$$

where we assume $1/s$ non-integer for convinience. 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 propability is equal to 1, which we get as an asymptotic result here. The result for $s > 1$ is somewhat more worrying. It says that Γ_n defined as sum over the instabilities of the periodic orbits 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.

16.2.3 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 stabilty of periodic orbits of the system, that is

$$|\mathcal{M}_n| = \Lambda_{10^{n-1}}^{-1} .$$

There is thus no problem in replacing the survival probability $\hat{\Gamma}_n$ given by (1.2), (14.2), that is the fraction of phase space $\mathcal{M}$ surviving n iterations of the map,

$$\hat{\Gamma}_n = \frac{1}{|\mathcal{M}|} \sum_i^{(n)} |\mathcal{M}_i| \,.$$

by a sum over periodic orbits of the form (**??**). The only orbit to worry about is the marginal fixed point $\overline{0}$ itself which we excluded from the zeta function (16.10).

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 probabilty in terms of cycle stabilties in chapter 14 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 fig. 16.2).

The generating partition is composed of the domains $\mathcal{M}_s$. The nth level partition $\mathcal{C}^{(n)} = \{\mathcal{M}_i\}$ can be constructed iteratively. Here i are words of length $n_i = n$, that is $i = s_2 s_2 \dots s_n$, and the intervals $\mathcal{M}_i$ are constructed iteratively

$$\mathcal{M}_{sj} = f_s^{-1}(\mathcal{M}_j) \quad , \tag{16.26}$$

where sj is the concatenation of letter s with word j of length $n_j < n$.

We will in the following concentrate on the survival probability $\hat{\Gamma}_n$, and consider other quantities of interest, like averages, later. 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 (14.6), the stability Λ_i of the periodic orbit i within the $\mathcal{M}_i$th strip, and bound the survival probability by the periodic orbit sum in (14.7).

?? on p. ??

The bound (14.6) relies on hyperbolicity, and is 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.

However, bounding survival probability strip by strip is not what is required for establishing the bound (14.7). 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 can be written by averaging over each prime cycle p separately

$$\mathcal{C}_1 \frac{1}{|\Lambda_p|} < \frac{1}{n_p} \sum_{i \in p} \frac{|\mathcal{M}_i|}{|\mathcal{M}|} < \mathcal{C}_2 \frac{1}{|\Lambda_p|}, \tag{16.27}$$

where the word i represents a code of the periodic orbit p and all its cyclic permutations. It can be shown that if one can find positive constants $\mathcal{C}_1$, $\mathcal{C}_2$ independent of p. Summing over all periodic orbits leads then again to (14.7). 8.6 on p. 165

To study averages of multiplicative weights we follow sect. 6.1 and introduce a phase space observable $a(x)$ and the integrated quantity

$$A^n(x) = \sum_{k=0}^{n-1} a(f^k(x)).$$

This leads us to introduce the generating function (6.10)

$$\langle e^{\beta\, A^n(x)} \rangle,$$

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 $\mathcal{C}_1$, $\mathcal{C}_2$ exist, such that

$$\mathcal{C}_1 \frac{e^{\beta A_p}}{|\Lambda_p|} < \frac{1}{n_p} \sum_{i \in p} \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}_Q} e^{\beta A^n(x)} dx < \mathcal{C}_2 \frac{e^{\beta A_p}}{|\Lambda_p|}, \tag{16.28}$$

is valid for all p. After performing the above average one gets

$$\mathcal{C}_1 \Gamma_n(\beta) < \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} e^{\beta A(x,n)} dx < \mathcal{C}_2 \Gamma_n(\beta), \tag{16.29}$$

with

$$\Gamma_n(\beta) = \sum_{p}^{n} \frac{e^{\beta A_p}}{|\Lambda_p|}. \tag{16.30}$$

and a dynamical zeta function can be derived. In the intermittent case one can expect that the bound (16.28) holds using an averaging argument similar to the

Figure 16.8: Markov graph corresponding to the alphabet $\{0^{k-1}1; \overline{0},\ k \geq 1\}$

one discussed in (16.27). 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 fig. 16.8.

☞ chapter ??

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 9

In the intermittent case, the fixed point $\overline{0}$ does not provide any shadowing (cf. sect. I.1), 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 (8.12), 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 expan-

sion 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 marignal $\overline{0}$ cycle is a dire necessity. The cycle weight $\Lambda_0^n = 1$ overestimates the corresponding interval length of $\mathcal{M}_{0^n}$ in the partition of the 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 $\mathcal{M}_{0^n}$ contribution), we are guaranteed to get at least the right asymptotical behaviour.

Note also, that if we are working with the spectral determinant (8.3), given in product form as

$$\det(1-z\mathcal{L}) = \prod_p \prod_{m=0}^{\infty} \left(1 - \frac{z^{n_p}}{|\Lambda_p|\Lambda_p^m}\right) \quad ,$$

for intermittent maps the marginal stable cycle has to be excluded. It intrudes an (unphysical) essential singularity at $z=1$ due the presence of a factor $(1-z)^\infty$ stemming from the $\overline{0}$ cycle.

16.3 General intermittent maps

Admittedly, the toy map is what is says - a toy model. The piece wise linearity of the map led to exact cancelations 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 choosen 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 technicallities - 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 mariginal 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, Fig. 16.3, or the repeller depicted in Fig. 16.2. The necessary modifications for multi-branch maps will briefly be discussed in the remark **??**. We again assume that the behaviour near the fixed point is given by Eq. (16.1). 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

$$\frac{1}{\Lambda_{10^n}} \sim \frac{1}{n^{1+1/s}} \ ,$$

∞ – alphabet		binary alphabet				
		$n=1$	$n=2$	$n=3$	$n=4$	$n=5$
1-cycle	n	1	10	100	1000	10000
2-cycle	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-cycle	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

Table 16.1: Infinite alphabet versus the original binary alphabet for the shortest periodic orbit families.

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 16.1. The topological length of the orbit is thus no longer determined by the iterations of our two-branch map, but by the number of times the cycle goes from the right to the left branch. Equivalently, one may define a new map, for which all the iterations on the left branch are done in one step. Such a map is called an *induced map* and the topological length of orbits in the infinite alphabet corresponds to the iterations of this induced map, see also remark **??**.

For generic intermittent maps, curvature contributions in the cycle expanded zeta function will not vanish exactly. The most natural way to organise 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

$$\begin{aligned}
\zeta^{-1} &= \prod_{p\neq 0}(1-t_p) = 1 - \sum_{n=1}^{\infty} c_n \qquad (16.31)\\
&= 1 - \sum_{n=1}^{\infty} t_n - \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \frac{1}{2}(t_{mn} - t_m t_n)\\
&- \sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}(\frac{1}{3}t_{kmn} - \frac{1}{2}t_{km}t^1_n + \frac{1}{6}t^1_k t^1_m t^1_n) - \sum_{l=1}^{\infty}\sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\ldots\ .
\end{aligned}$$

The first sum is the fundamental term, which we have already seen in the toy model, cf. Eq. 16.10. The curvature terms c_n in the expansion are now n-fold infinite sums where the prefactors take care of double counting of primitiv periodic orbits. Note that diagonal contributions of the form $t_{nn} = (t_n)^2$ cancel exactly, as they should.

Let's 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

$$f(z) = \sum_{n=0}^{\infty} c_n z^n. \qquad (16.32)$$

with algebraically decreasing coefficients

$$c_n \sim \frac{1}{n^\alpha} \quad \text{with} \quad \alpha > 0$$

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 branchcuts of the zeta function are to be expected. By carefully subtracting the asymptotic behaviour with the help of Eq. (16.11) or Eq. (16.12), one can in general construct an analytic continuation of $f(z)$ around $z = 1$ of the form

$$\begin{aligned}
f(z) &\sim A(z) + (1-z)^{\alpha-1}B(z) \qquad \alpha \notin \mathbb{N} \qquad (16.33)\\
f(z) &\sim A(z) + (1-z)^{\alpha-1}\ln(1-z) \qquad \alpha \in \mathbb{N},
\end{aligned}$$

where $A(z)$ and $B(z)$ are functions analytic in a disc around $z = 1$. We thus again find that the zeta function (16.31) 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 behaviour derived in Eqs. (16.24) and (16.25) is a general result, that is, the survival propability is given asymptotically by

$$\Gamma_n \sim C\frac{1}{n^{1/s}} \tag{16.34}$$

for all 1-dimensional maps of the form (**??**). 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 burried 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 two different strategies which we will sketch next.

16.3.1 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 (16.33) at $z = 1$. This can be done in principle, if the coefficients c_n in sums like (16.32) are known (as for our toy model). One then considers a resummation of the form

$$\sum_{j=0}^{\infty} c_j z^j = \sum_{j=0}^{\infty} a_j (1-z)^j + (1-z)^{\alpha-1} \sum_{j=0}^{\infty} b_j (1-z)^j i, \tag{16.35}$$

and the coefficients a_j and b_j are obtained in terms of the c_j's by expanding $(1-z)^j$ and $(1-z)^{j+\alpha-1}$ on the right hand side around $z = 0$ using (16.11) and equating the coefficients.

In practical calculations one often has only a finite number of coefficients c_j, $0 \le j \le n_N$, 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 (16.35). We replace the infinite sums in (16.35) by finite sums of increasing degrees n_a and n_b, and require that

$$\sum_{i=0}^{n_a} a_i (1-z)^i + (1-z)^{\alpha-1} \sum_{i=0}^{n_b} b_i (1-z)^i = \sum_{i=0}^{n_N} c_i z^i + O(z^{n_N+1}) \quad . \tag{16.36}$$

One proceeds again by expanding the right hand side around $z = 0$, skipping all powers z^{n_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 (16.36)

are adjacent. If one chooses $n_a+n_b+2 = n_N+1$, then, for each cutoff length n_N, the integers n_a and n_b are uniquely determined from a linear system of equations. The prize we pay is that the so obtained coefficients depend on the cutoff n_N. One can now study convergence of the coefficients a_j, and b_j, with respect to increasing values of n_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 (16.34), cf. (16.23). The resummed expression can also be used to compute zeros, inside or outside the radius of convergence of the cycle expansion $\sum c_j z^j$. The scheme outlined in this section tacitly assumes that a representation of form (16.33) holds in a disc of radius 1 around $z = 1$. Any additional knowledge of the asymptotics of sums like (16.32) can of course be built into the ansatz (16.35) improving the convergence.

16.3.2 Analytical continuation by integral transformations

In this section, we will introduce a method wich provides an analytic continuation of sums of the form (16.32) without explicitly relying on an ansatz (16.35). The main idea is to rewrite the sum (16.32) 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 $c_n \equiv c(n)$ explicitely for all n. If the coefficients are not known analytically, one may procede by approximating the large n behaviour in the form

$$c(n) = n^{-\alpha}(C_1 + C_2 n^{-1} + \ldots)$$

and determine the constants C_i numerically from periodic orbit data. By using the identity

$$\sum_{n=-\infty}^{\infty} \delta(x-n) = \sum_{m=-\infty}^{\infty} \exp(2\pi \mathrm{i} m x) \tag{16.37}$$

obtained by Poisson summation, we may write the sum (16.32)

$$f(z) = \frac{1}{2}c(0) + \sum_{m=-\infty}^{\infty} \int_0^{\infty} dx\, e^{2\pi \mathrm{i} m x} c(x) z^x. \tag{16.38}$$

The continuous variable x corresponds to the discrete summation index n and it is convenient to write $z = r\exp(\mathrm{i}\sigma)$ from now on. The integrals are of course still not convergent for $r > 0$. 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 form the real onto the imaginary axis. For the $m = 0$ term in (16.38), we transform $x \to \mathrm{i}x$ and the integral takes on the form

$$\int_0^\infty dx\, c(x)\, r^x\, e^{\mathrm{i}x\sigma} = \mathrm{i}\int_0^\infty dx\, c(ix)\, r^{\mathrm{i}x} e^{-x\sigma}.$$

The integrand is now exponentially decreasing for all $r > 0$ and $\sigma \neq 0$ or 2π. The last condition reminds us again of the existence of a branch cut at $\mathrm{Re}\, z \geq 1$. By the same technique, we find the analytic continuation for all the other integrals in (16.38). The real axis is then rotated according to $x \to \mathrm{sign}(m)\mathrm{i}x$ where $\mathrm{sign}(m)$ refers to the sign of m.

$$\int_0^\infty dx\, e^{\pm 2\pi \mathrm{i}|m|x} c(x)\, r^x e^{\mathrm{i}x\sigma} = \pm\mathrm{i}\int_0^\infty dx\, c(\pm\mathrm{i}x)\, r^{\pm\mathrm{i}x} e^{-x(2\pi|m|\pm\sigma)}.$$

Changing summation and integration, we can carry out the sum over $|m|$ explicitely and one finally obtains the compact expression

$$\begin{aligned} f(z) &= \frac{1}{2}c(0) + \mathrm{i}\int_0^\infty dx\, c(ix)\, r^{\mathrm{i}x} e^{-x\sigma} \\ &+ \mathrm{i}\int_0^\infty dx\, \frac{e^{-2\pi x}}{1-e^{-2\pi x}} \left[c(\mathrm{i}x) r^{\mathrm{i}x} e^{-x\sigma} - c(-\mathrm{i}x) r^{-\mathrm{i}x} e^{x\sigma}\right]. \end{aligned} \tag{16.39}$$

The transformation from the original sum to the two integrals in (16.39) is exact for $r \leq 1$, and provides an analytic continuation for $r > 0$. The expression (16.39) is especially useful for an efficient numerical calculations of the dynamical zeta function for $|z| > 1$, which is essential when searching for zeros and poles of an dynamical zeta function.

16.3.3 Curvature contributions

So far, we have discussed the fundamental term $\sum_{n=1}^\infty t_n$ in (16.31), and showed ways to extract the important information from 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 (16.33) are contained in the curvature terms in (16.31). These are multiple sums of the form

$$\sum_{m=1}^\infty \sum_{n=1}^\infty \frac{1}{2}(t_{mn} - t_m t_n)$$

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 cycle expansion itself suggests that the terms t_{mn} decrease asymptotically like

$$t_{mn} \sim \frac{1}{(nm)^{1+1/s}} \tag{16.40}$$

for 2-cycles of the induced map and in general for n-cycles like

$$t_{m_1 m_2 \ldots m_n} \sim \frac{1}{(m_1 m_2 \ldots m_n)^{1+1/s}}.$$

If we know the cycle weights $t_{m_1 m_2 \ldots m_n}$ analytically, we may proceed like in sect. 16.3.2, transform the multiple sums into multiple integrals and rotate the axis' of integration. Otherwise, we have to look for clever ways to approximate, for example, the weights t_{mn} in terms of 2-dimensional functions with asymptotics (16.40).

We are certainly reaching the borderline of what can be done and is worth the effort to actually compute zeta functions from periodic orbit data. In the next section, we will introduce a method applicable for intermittent maps, which does not rely on periodic orbit data, but is based on a probabilistic approch. This circumvents some of the technical problems encountered in this section.

16.4 Probabilistic or BER zeta functions

So far we discussed 1-dimensional model systems in order to investigate dynamical implications of marginal fixed points. We describe next how probabilistic methods may be employed in order to write down approximate dynamical zeta functions for intermittent systems.

We will discuss the method in the most general setting, that is for a flow in arbitrary dimensions. The key idea is to introduce a surface of section $\mathcal{P}$ such that all trajectories fromving from this section will have spent some time near the marginal stable fixed point and in the chaotic phase. An important quantitiy in what follows is the time of *first return function* $T(x)$ which gives the the time of flight of a trajectory starting in x to return to the surface of section $\mathcal{P}$. It depends

only on the phase space coordinate x on the surface of section. The period (3.2) of a periodic orbit p intersecting the section $\mathcal{P}$ n_p times can be written

$$T_p = \sum_{n=0}^{n_p-1} T(f^n(x_p)),$$

where $f(x)$ is the Poincaré map, and x_p is a cycle point, a point where p intersects the section. Next we formulate a dynamical zeta function associated with the cycle weights

$$e^{\beta A_p - sT_p}$$

where

$$A_p = \sum_{n=0}^{n_p-1} a(f^n(x_p)),$$

as

$$1/\zeta(z,s,\beta) = \prod_p \left(1 - \frac{z^{n_p} e^{\beta A_p - sT_p}}{|\Lambda_p|}\right) \quad . \tag{16.41}$$

The zeta function captures the dynamics of both the flow *and* the Poincaré map. The dynamical zeta function for the flow is obtained as $1/\zeta(s,\beta) = 1/\zeta(z = 1, s, \beta)$. and the dynamical zeta function of the Poincaré map is $1/\zeta(z,\beta) = 1/\zeta(z, s = 0, \beta)$. If we are dealing with a discrete dynamical system given by a map instead of a flow, we can use the same formalism by setting $T_p = n_p$ and $z = e^{-s}$.

Our basic approximation will be a *stochasticity assumption*. 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 $\mathcal{P}$. Under the above stochasticity assumption the large n behavior is

$$\langle e^{\beta A(x_0,n) - sT(x_0,n)} \rangle_{\mathcal{P}} \sim \left(\int_{\mathcal{P}} e^{\beta a(x) - sT(x)} \rho(x) dx \right)^n , \tag{16.42}$$

where $\rho(x)$ is the invariant density of the Poincaré map. This type of behavior is equivalent to there being only one zero $z_0(s,\beta) = \int e^{\beta a(x) - st(xt)} \rho(x) dx$ of

$1/\zeta(z,s,\beta)$ in the $z-\beta$ plane. In the language of Ruelle 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

$$1/\zeta(z,s,\beta) = z - \int_{\mathcal{P}} e^{\beta a(x) - sT(x)} \rho(x) dx \quad , \tag{16.43}$$

where we have neglected a possible analytic and non-zero prefactor. The dynamical zeta function of the flow is now

$$1/\zeta(s,\beta) = 1/\zeta(z=1,s,\beta) = 1 - \int e^{-sT(x)} e^{\beta a(x)} \rho(x) dx \quad . \tag{16.44}$$

Normally, the best one can hope for is a finite gap to the leading resonance of the Poincaré map. The above dynamical zeta function is then only approximatively valid. As it is derived from an approximation due to Baladi–Eckmann-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

$$\psi(T) = \int \delta(T - T(x)) \rho(x) dx \tag{16.45}$$

16.7 on p. 380

The BER zeta function at $\beta = 0$ is then given in terms of the Laplace transform of this distribution

$$1/\zeta_{\rm BER}(s) = 1 - \int_0^\infty \psi(T) e^{-sT} dT.$$

16.8 on p. 380

Example 1

For the binary shift map

$$x \mapsto f(x) = 2x \bmod 1, \tag{16.46}$$

one easily derives the distribution of return times

$$\psi_n = \frac{1}{2^n} \qquad n \geq 1. \tag{16.47}$$

The BER zeta function becomes

$$1/\zeta_{\rm BER}(z) = 1 - \sum_{n=1}^{\infty} \psi_n z^n = 1 - \sum_{n=1}^{\infty} \frac{z^n}{2^n} = \frac{1-z}{1-z/2} \equiv \zeta^{-1}(z)/(1-z/\Lambda_0).$$

That is, the "approximate" zeta function is in this case the exact dynamical zeta function, with the cycle point $\overline{0}$ pruned.

Example 2

For the toy model presented in sect. 16.2.1 one gets

$$\psi_1 = |\mathcal{M}_1|, \tag{16.48}$$

and

$$\psi_n = |\mathcal{M}_n| \frac{1-b}{1-a},$$

for $n \geq 2$ leading to a BER zeta function

$$1/\zeta_{\rm BER}(z) = 1 - z|\mathcal{M}_1| - \sum_{n=2}^{\infty} |pS_n| z^n,$$

which coincides again with the exact result, see refeqcyc-exp-toy.

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 a complete Markov partitions. The stochasticity assumption on which the probabilistic approximation is based is fulfilled here exactly. Curvature terms of in a cycle expansion are as a consequence identically zero. Introducing a small nonlinearity will change matters completely. The BER zeta function and the fundamental part of a cycle expansion as discussed in sect. 13.1.1 are indeed intricately related. Note, however, that they are not identical in general; the BER zeta function obeys the flow conservation sum rule by construction, whereas the fundamental part of a cycle expansion does not.

Commentary

Remark **16.1** What about the evolution operator formalism? The main motivation for introducing evolution operators was their semigroup property (6.21). This was very natural for hyperbolic systems where instabilites grow exponentially, and evolution operators capture this behaviour 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 althogether. 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 implicitely by changing from a finite to an infinite alphabet. We refer to refs. [3, 4] for a detailed discussion (and applications to 1-dimensional maps) of this technique.

Remark **16.2** Intermittency. Intermittency in the Lorenz system was discovered by Manneville and Pomeau [1]. Manneville and Pomeau demonstrated that in neighborhood of parameter value $r_c = 166.07$ the mean duration of the periodic motion scales just as for 1-d maps, as $(r - r_c)^{1/2}$. In ref. [5] they explained this phenomenon in terms of a 1-dimensional map such as (16.1) near tangent bifurcation, and classified possible types of intermittency.

Farey map (13.26) has been studied widely in the context of intermittent dynamics, for example in refs. [16, 17, 2, 18, 23, 13, 2]. For the Farey map (13.26) and the related Gauss shift map (19.38) (which is the induced map of the Farey map), the Fredholm determinant and the dynamical zeta functions 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 [**?**] for "Axiom A" hyperbolic systems.

Piecewise linear models like the one considered here have been studied by Gaspard and Wang [6]. The escape problem has been treated along the way employed here in ref. [7], resummations have been considered in ref. [8]. The proof of the bound (16.27) can be found in Dahlqvist's unpublished notes `www.nbi.dk/ChaosBook/extras/PDahlqvistEscape.ps.gz`.

Remark **16.3** Jonquière functions. In statistical mechanics Jonquière functions appear in the theory of free Bose-Einstein gas, see refs.??

Remark **16.4** Tauberian theorems. In this chapter we used Tauberian theorems for power series and Laplace transforms: a highly recommended reading on this issue is ref. [9].

Remark **16.5** Probabilistic methods, BER zeta functions. The use of probabilistic methods in intermittent chaos is widespread [10]. The BER approximation studied here is inspired by Baladi, Eckmann and Ruelle [14] and has been developped in refs. [13, 15].

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 initially developped for fully hyperbolic systems. 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 towards the marginally stable cycles. Algebraically decaying correlations are controlled by the branch cuts in the dynamical zeta functions. Compared to pure hyperbolic systems, the physical consequences are drastic: exponential decays are replaced by slowly decaying power laws, and diffusion becomes anomalous.

References

[16.1] P. Manneville and Y. Pomeau, *Phys. Lett.* **75A**, 1 (1979).

[16.2] H.H. Rugh, *Inv. Math.* **135**, 1 (1999).

[16.3] 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, *J. Stat. Phys.* **66**, 503 (1992).

[16.4] S. Isola, *J. Stat. Phys.* **97**, 263 (1999).

[16.5] Y. Pomeau and P. Manneville, *Commun. Math. Phys.* **74**, 189 (1980).

[16.6] 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).

[16.7] P. Dahlqvist, *Phys. Rev. E* **60**, 6639 (1999).

[16.8] P. Dahlqvist, *J. Phys. A* **30**, L351 (1997).

[16.9] W. Feller, *An introduction to probability theory and applications, Vol. II* (Wiley, New York 1966).

[16.10] T. Geisel and S. Thomae, *Phys. Rev. Lett.* **52**, 1936 (1984).

[16.11] T. Geisel, J. Nierwetberg and A. Zacherl, *Phys. Rev. Lett.* **54**, 616 (1985).

[16.12] R. Artuso, G. Casati and R. Lombardi, *Phys. Rev. Lett.* **71**, 62 (1993).

[16.13] P. Dahlqvist, *Nonlinearity* **8**, 11 (1995).

[16.14] V. Baladi, J.-P. Eckmann and D. Ruelle, *Nonlinearity* **2**, 119 (1989).

[16.15] P. Dahlqvist, *J. Phys. A* **27**, 763 (1994).

[16.16] D.H. Mayer, *Bull. Soc. Math. France* **104**, 195 (1976).

[16.17] D. Mayer and G. Roepstorff, *J. Stat. Phys.* **47**, 149 (1987).

[16.18] D. H. Mayer, *Continued fractions and related transformations*, in ref. [11].

[16.19] S. Isola, *Dynamical zeta functions for non–uniformly hyperbolic transformations*, Universitá degli Studi di Bologna preprint (1995).

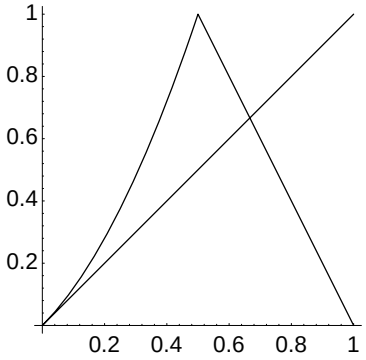
Exercises

16.1 Integral representation of Jonquière functions. Check the integral representation (**??**). Notice how the denominator is connected to Bose-Einstein distribution. Compute $J(x+i\epsilon) - J(x-i\epsilon)$ for a real $x > 1$.

16.2 Power law correction to a power law. Expand (16.20) further and derive the leading power law correction to (16.23).

16.3 Inverse Laplace. Consider (**??**) in the case of *discrete* time mappings: show that it can be rewritten in a form analogous to (**??**).

16.4 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

$$\Lambda_{\underbrace{00\cdots0}_{n}1} \sim n^2$$

16.5 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 (4.68) is that the Λ_n grows only as a power law in number of bounces:

$$\Lambda_n \propto n_k^2 \,. \tag{16.49}$$

Check.

16.6 Accelerated diffusion. Consider a map h, such that $\hat{h} = \hat{f}$ of fig. **??**, 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

$$\sigma^2(t) \sim \begin{cases} t & \text{for } \alpha > 2 \\ t \ln t & \text{for } \alpha = 2 \\ t^{3-\alpha} & \text{for } \alpha \in (1,2) \\ t^2/\ln t & \text{for } \alpha = 1 \\ t^2 & \text{for } \alpha \in (0,1) \end{cases}$$

16.7 Recurrence times for Lorentz gas with infinite horizon. Consider the Lorentz gas with unbounded horizon with a square lattice geometry, with disk radius R and unit lattice spacing. Label disks according to the (integer) coordinates of their center: the sequence of recurrence times $\{t_j\}$ is given by the set of collision times. Consider orbits that leave the disk sitting at the origin and hit a disk far away after a free flight (along the horizontal corridor). Initial conditions are characterized by coordinates (ϕ, α) (ϕ determines the initial position along the disk, while α gives the angle of the initial velocity with respect to the outward normal: the appropriate measure is then $d\phi \cos\alpha \, d\alpha$ ($\phi \in [0, 2\pi)$, $\alpha \in [-\pi/2, \pi/2]$. Find how $\psi(T)$ scales for large values of T: this is equivalent to investigating the scaling of portions of the phase space that lead to a first collision with disk $(n, 1)$, for large values of n (as $n \mapsto \infty$ $n \simeq T$).

Suggested steps

(a) Show that the condition assuring that a trajectory indexed by (ϕ, α) hits the (m, n) disk (all other disks being transparent) is written as

$$\left| \frac{d_{m,n}}{R} \sin\left(\phi - \alpha - \theta_{m,n}\right) + \sin\alpha \right| \leq 1 \tag{16.50}$$

where $d_{m,n} = \sqrt{m^2 + n^2}$ and $\theta_{m,n} = \arctan(n/m)$. You can then use a small R expansion of (16.50).

(b) Now call j_n the portion of the phase space leading to a first collision with disk $(n,1)$ (take into account screening by disks $(1,0)$ or $(n-1,1)$). Denote by $J_n = \bigcup_{k=n+1}^{\infty} j_k$ and show that $J_n \sim 1/n^2$, from which the result for the distribution function follows.

16.8 Probabilistic zeta function for maps. Derive the probabilistic zeta function for a map with recurrence distribution ψ_n.

Chapter 17

Discrete symmetries

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 in quantum mechanics. In the process we also learn how to simplify the classical dynamics. 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_α-dimensional irreducible representations D_α of the symmetry group,

$$(1-t_p)^{m_p} = \prod_\alpha \det\, (1 - D_\alpha(h_{\tilde{p}})t_{\tilde{p}})^{d_\alpha}\ , \quad t_p = t_{\tilde{p}}^{g/m_p}\ ,$$

where $t_{\tilde{p}}$ is the cycle weight evaluated on the fundamental domain, g is the dimension 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 the dynamical zeta functions have particularly simple cycle expansions, a simple geometrical shadowing interpretation of their convergence, and as they suffice for determination of leading eigenvalues, we shall concentrate in this chapter on their factorizations; the full spectral determinants can be factorized by the same techniques. To emphasize the group theoretic structure of zeta functions, we shall combine all the non-group-theory dependence of a p-cycle into a cycle weight t_p.

This chapter is meant to serve as a detailed guide to 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 out a series of cases of physical

interest: C_2, C_{3v} symmetries in this chapter, and C_{2v}, C_{4v} symmetries in appendix H below.

17.1 Preview

Dynamical systems often come equipped with discrete symmetries, such as the reflection and the rotation symmetries of various potentials. Such symmetries simplify and improve the cycle expansions in a rather beautiful way; they can be exploited to relate classes of periodic orbits and reduce dynamics to a fundamental domain. Furthermore, 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.

Invariance of a system under symmetries means that the symmetry image of a cycle is again a cycle, with the same weight. 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, shifted in time. A cycle is *symmetric* if some symmetry operations act on it like a shift in time, advancing the starting point to the starting point of a symmetry related segment. A symmetric cycle can thus be subdivided into a sequence of repeats of an *irreducible segment*. The period or any average evaluated along the full orbit is given by the sum over the segments, whereas the stability is given by the product of the stability matrices of the individual segments.

Cycle degeneracies induced by the symmetry are removed by *desymmetrization*, reduction of the full dynamics to the dynamics on a *fundamental domain*. The phase space can be completely tiled by a fundamental domain and its symmetry images. The irreducible segments of cycles in the full space, folded back into the fundamental domain, are closed orbits in the reduced space.

17.1.1 3-disk game of pinball

We have already exploited a discrete symmetry in our introduction to the 3-disk game of pinball, sect. 1.3. As the three disks are equidistantly spaced, our game of pinball has a sixfold symmetry. The symmetry group of relabelling the 3 disks is the permutation group S_3; however, it is better to think of this group geometrically, as C_{3v}, the group of rotations by $\pm 2\pi/3$ and reflections across the three symmetry axes. Applying an element (identity, rotation by $\pm 2\pi/3$, or one of the three possible reflections) of this symmetry group to any trajectory yields another trajectory. 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 relabelling of the disks.

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 fig. 10.4. 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 degeneracy 2, reflectionally symmetric ones (such as the 2-cycles $\overline{12}$, $\overline{13}$ and $\overline{23}$) have degeneracy 3, and global orbits with no symmetry are 6-fold degenerate. Table 10.2 lists some of the shortest binary symbols strings, together with the corresponding full 3-disk symbol sequences and orbit symmetries. Some examples of such orbits are shown in fig. 1.4.

We shall return to the 3-disk game of pinball desymmetrization in sects. 17.2.2 and 17.6, but first we develop a feeling for discrete symmetries by working out a simple 1-d example.

17.1.2 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 fig. 17.1. Denote the reflection operation by $\mathbf{C}x = -x$. The symmetry of the map implies that if $\{x_n\}$ is a trajectory, than also $\{\mathbf{C}x_n\}$ is a trajectory because $\mathbf{C}x_{n+1} = \mathbf{C}f(x_n) = f(\mathbf{C}x_n)$. The dynamics can be restricted to a fundamental domain, in this case to one half of the original interval; every time a trajectory leaves this interval, it can be mapped back using $\mathbf{C}$. Furthermore, the evolution operator commutes with $\mathbf{C}$, $\mathcal{L}(y,x) = \mathcal{L}(\mathbf{C}y, \mathbf{C}x)$. $\mathbf{C}$ satisfies $\mathbf{C}^2 = \mathbf{e}$ and can be used to decompose the phase space into mutually orthogonal symmetric and antisymmetric subspaces by means of projection operators

$$\begin{aligned} P_{A_1} &= \frac{1}{2}(\mathbf{e}+\mathbf{C}) \ , \qquad P_{A_2} = \frac{1}{2}(\mathbf{e}-\mathbf{C}) \, , \\ \mathcal{L}_{A_1}(y,x) &= P_{A_1}\mathcal{L}(y,x) = \frac{1}{2}\left(\mathcal{L}(y,x)+\mathcal{L}(-y,x)\right) \ , \\ \mathcal{L}_{A_2}(y,x) &= P_{A_2}\mathcal{L}(y,x) = \frac{1}{2}\left(\mathcal{L}(y,x)-\mathcal{L}(-y,x)\right) \ . \end{aligned} \tag{17.1}$$

To compute the traces of the symmetrization and antisymmetrization projection operators (17.1), 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: $\det(1-\mathcal{L}) = \det(1-\mathcal{L})_a \det(1-\mathcal{L})_{\tilde{s}} \det(1-\mathcal{L})_b$.

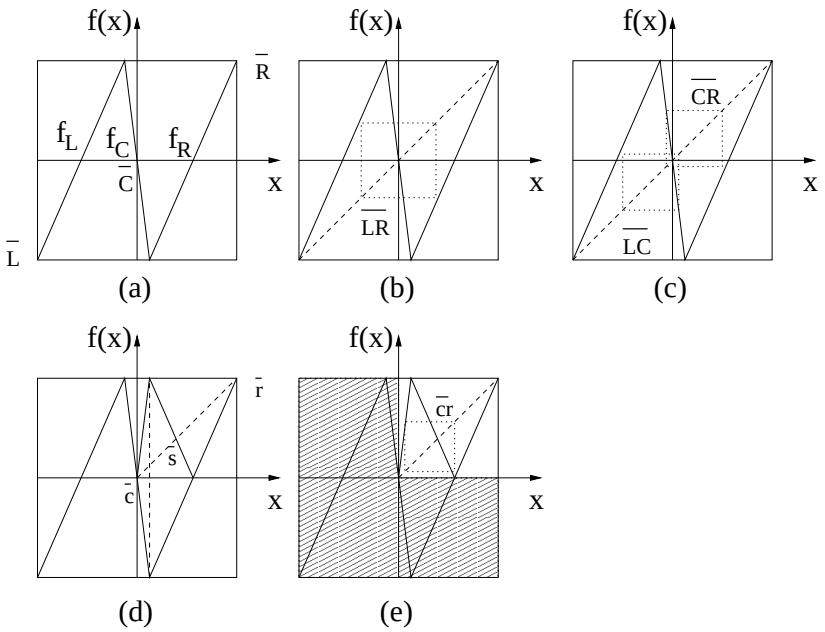


Figure 17.1: The Ulam sawtooth map with the C_2 symmetry $f(-x) = -f(x)$. (a) boundary fixed point $\overline{C}$, (b) symmetric 2-cycle $\overline{LR}$, (c) asymmetric 2-cycles pair $\{\overline{LC},\overline{CR}\}$. The Ulam sawtooth map restricted to the fundamental domain; pieces of the global map (a) are reflected into the upper right quadrant. (d) Boundary fixed point $\overline{C}$ maps into the fixed point $\overline{c}$, symmetric 2-cycle $\overline{LR}$ maps into fixed point $\overline{s}$, and the asymmetric fixed point pair $\{\overline{L},\overline{R}\}$ maps into a single fixed point $\overline{r}$, (e) the asymmetric 2-cycles pair $\{\overline{LC},\overline{CR}\}$ maps into a single 2-cycle $\overline{cr}$.

Asymmetric cycles: A periodic orbits is not symmetric if $\{x_a\} \cap \{\mathbf{C}x_a\} = \emptyset$, where $\{x_a\}$ is the set of periodic points belonging to the cycle a. Thus $\mathbf{C}$ 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 (17.1); as they are degenerate, the prefactor $1/2$ cancels. Resumming as in the derivation of (8.12) 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 $\mathbf{C}$ 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 $n_{\tilde{s}}$, $f^{n_{\tilde{s}}}(x_s) = \mathbf{C}x_s$. $\delta(x - f^n(x))$ picks up $2n_{\tilde{s}}$ contributions for every even traversal, $n = rn_{\tilde{s}}$, r even, and $\delta(x + f^n(x))$ for every odd traversal, $n = rn_{\tilde{s}}$, r odd. Absorb the group-theoretic prefactor in the 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 $\mathcal{M}_s$ 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 resumming 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 $\delta(x + f^n(x))$:

$$\begin{aligned} z^n \mathrm{tr}\, \mathcal{L}_{\pm}^n &\to \int_{\mathcal{M}_b} dx\, z^n\, \frac{1}{2}\, (\delta(x - f^n(x)) \pm \delta(x + f^n(x))) \\ &= \sum_{r=1}^{\infty} \delta_{n,r}\, t_b^r\, \frac{1}{2} \left(\frac{1}{1 - 1/\Lambda_b^r} \pm \frac{1}{1 + 1/\Lambda_b^r} \right) \\ z^n\, \mathrm{tr}\, \mathcal{L}_{+}^n &\to \sum_{r=1}^{\infty} \delta_{n,r} \frac{t_b^r}{1 - 1/\Lambda_b^{2r}}\,; \qquad z^n\, \mathrm{tr}\, \mathcal{L}_{-}^n \to \sum_{r=1}^{\infty} \delta_{n,r} \frac{1}{\Lambda_b^r} \frac{t_b^r}{1 - 1/\Lambda_b^{2r}}\,. \end{aligned}$$

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 even derivatives contribute to the symmetric subspace (and odd to the antisymmetric subspace) because the orbit lies on the boundary.

Finally, the symmetry reduced spectral determinants follow by collecting the above results:

$$F_+(z) = \prod_a \prod_{k=0}^{\infty} \left(1 - \frac{t_a}{\Lambda_a^k}\right) \prod_{\tilde{s}} \prod_{k=0}^{\infty} \left(1 - \frac{t_{\tilde{s}}}{\Lambda_{\tilde{s}}^k}\right) \prod_{k=0}^{\infty} \left(1 - \frac{t_b}{\Lambda_b^{2k}}\right)$$

$$F_-(z) = \prod_a \prod_{k=0}^{\infty} \left(1 - \frac{t_a}{\Lambda_a^k}\right) \prod_{\tilde{s}} \prod_{k=0}^{\infty} \left(1 + \frac{t_{\tilde{s}}}{\Lambda_{\tilde{s}}^k}\right) \prod_{k=0}^{\infty} \left(1 - \frac{t_b}{\Lambda_b^{2k+1}}\right) \tag{17.2}$$

We shall work out the symbolic dynamics of such reflection symmetric systems in some detail in sect. 17.5. 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.

17.1 on p. 403

17.2 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 (**??**) this means

$$\begin{aligned} f(x) &= \mathbf{g}^{-1} f(\mathbf{g}x) \\ \mathcal{L}(y,x) &= \mathcal{L}(\mathbf{g}y, \mathbf{g}x) \ . \end{aligned} \tag{17.3}$$

Bold face letters for group elements indicate a suitable representation on phase 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 π around the origin. The map f must then commute with rotations by π, $f(\mathbf{C}x) = \mathbf{C}f(x)$, with $\mathbf{C}$ given by the $[2 \times 2]$ matrix

$$\mathbf{C} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{17.4}$$

$\mathbf{C}$ satisfies $\mathbf{C}^2 = e$ and can be used to decompose the phase space into mutually orthogonal symmetric and antisymmetric subspaces by means of projection operators (17.1). More generally the projection operator onto the α irreducible subspace of dimension d_α is given by $P_\alpha = (d_\alpha/|G|) \sum \chi_\alpha(h) \mathbf{h}^{-1}$, where $\chi_\alpha(h) = \operatorname{tr} D_\alpha(h)$ are the group characters, and the transfer operator $\mathcal{L}$ splits into a sum of inequivalent irreducible subspace contributions $\sum_\alpha \operatorname{tr} \mathcal{L}_\alpha$,

$$\mathcal{L}_\alpha(y,x) = \frac{d_\alpha}{|G|} \sum_{h \in G} \chi_\alpha(h) \mathcal{L}(\mathbf{h}^{-1}y, x) \ . \tag{17.5}$$

The prefactor d_α in the above reflects the fact that a d_α-dimensional representation occurs d_α times.

17.2.1 Cycle degeneracies

If $g \in G$ is a symmetry of the dynamical problem, the weight of a cycle p and the weight of its image under a symmetry transformation g are equal, $t_{gp} = t_p$. The number of degenerate cycles (topologically distinct, but mapped into each other by symmetry transformations) depends on the cycle symmetries. Associated with a given cycle p is a maximal subgroup $\mathcal{H}_p \subseteq G$, $\mathcal{H}_p = \{e, b_2, b_3, \ldots, b_h\}$ of order h_p, whose elements leave p invariant. The elements of the quotient space $b \in G/\mathcal{H}_p$ generate the degenerate cycles bp, so the multiplicity of a degenerate cycle is $m_p = g/h_p$.

Taking into account these degeneracies, the Euler product (8.12) takes the form

$$\prod_p (1 - t_p) = \prod_{\hat{p}} (1 - t_{\hat{p}})^{m_{\hat{p}}} . \tag{17.6}$$

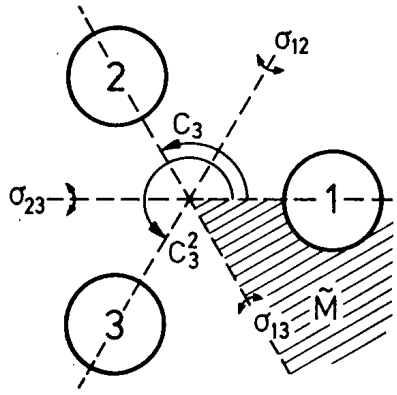


Figure 17.2: The symmetries of three disks on an equilateral triangle. The fundamental domain is indicated by the shaded wedge.

Here $\hat{p}$ is one of the m_p degenerate cycles, picked to serve as the label for the entire class. Our labelling convention is usually lexical, *i.e.*, we label a cycle p by the cycle point whose label has the lowest value, and we label a class of degenerate cycles by the one with the lowest label $\hat{p}$. In what follows we shall drop the hat in $\hat{p}$ when it is clear from the context that we are dealing with symmetry distinct classes of cycles.

17.2.2 Example: C_{3v} invariance

An illustration of the above is afforded by C_{3v}, the group of symmetries of a game of pinball with three equal size, equally spaced disks, fig. 17.2. The group consists 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 $\{C_3, C_3^2\}$, so its dimension is $g = 6$. On the disk labels $\{1, 2, 3\}$ these symmetries act as permutations which map cycles into cycles. For example, the flip across the symmetry axis going through disk 1 interchanges the symbols 2 and 3; it maps the cycle $\overline{12123}$ into $\overline{13132}$, fig. 1.4a.

The subgroups of C_{3v} are C_v, consisting of the identity and any one of the reflections, of dimension $h = 2$, and $C_3 = \{e, C_3, C_3^2\}$, of dimension $h = 3$, so possible cycle multiplicities are $g/h = 2$, 3 or 6.

The C_3 subgroup invariance is exemplified by the 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, fig. 1.4b; $\mathcal{H}_p = \{e, C_3, C_3^2\}$, and the degeneracy is $g/h_{c_3} = 2$.

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 $\mathcal{H}_{\hat{p}} = \{e, \sigma_{23}\}$. Its order is $h_{C_v} = 2$, so the degeneracy is $m_{\hat{p}} = g/h_{C_v} = 3$; the cycles in this class, $\overline{1213}$, $\overline{1232}$ and $\overline{1323}$, are related by $2\pi/3$ rotations, fig. 1.4(c).

A cycle of no symmetry, such as $\overline{12123}$, has $\mathcal{H}_p = \{e\}$ and contributes in all

six terms (the remaining cycles in the class are $\overline{12132}$, $\overline{12313}$, $\overline{12323}$, $\overline{13132}$ and $\overline{13232}$), fig. 1.4a.

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 are related by no space symmetry (fig. 1.4(d)).

The Euler product (8.12) for the C_{3v} symmetric 3-disk problem is given in (13.31).

17.3 Dynamics in the fundamental domain

So far we have used the discrete symmetry to effect a reduction in the number of independent cycles in cycle expansions. The next step achieves much more: the symmetries can be used to restrict all computations to a *fundamental domain*. We show here that to each global cycle p corresponds a fundamental domain 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 $h_{\tilde{p}}$.

An important effect of a discrete symmetry is that it tesselates the phase space into copies of a fundamental domain, and thus induces a natural partition of phase space. The group elements $g = \{a, b, \cdots, d\}$ which map the fundamental domain $\tilde{M}$ into its copies $g\tilde{M}$, can double in function as letters of a symbolic dynamics alphabet. 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. We demonstrate the reduction for a series of specific examples in sect. 17.4. An immediate gain arising from symmetry invariant relabelling is that N-disk symbolic dynamics becomes $(N-1)$-nary, with no restrictions on the admissible sequences. However, the main gain is in the close connection between the symbol string symmetries and the phase space symmetries which will aid us in the dynamical zeta function factorizations. Once the connection between the full space and the reduced space is established, working in the fundamental domain (*ie.*, with irreducible segments) is so much simpler that we never use the full space orbits in actual computations.

If the dynamics is invariant under a discrete symmetry, the phase 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 (17.4) with symmetry group $G = \{e, C\}$, the phase 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 π.

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 phase 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 phase-space coordinates. For instance, in the above example (17.4) $h \in C_2$ and $D(h)$ can be either the identity or the interchange of the two domain labels,

$$D(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} , \quad D(C) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} . \tag{17.7}$$

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 phase-space transformation $\mathbf{h} \neq \mathbf{e}$ leaves invariant sets of *boundary* points; for example, under reflection σ across a symmetry axis, the axis itself remains invariant. The boundary periodic orbits that belong to such point-wise invariant sets will require special care in $\mathrm{tr}\, \mathcal{L}$ evaluations.

One can associate to the evolution operator (**??**) 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 (17.3) 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 (8.3) required for the evaluation of the eigenvalues of the transfer operator can be evaluated on the fundamental domain alone

$$\mathrm{tr}\,\mathcal{L} = \int_M dx \mathcal{L}(x,x) = \int_{\tilde{M}} d\tilde{x} \sum_h \mathrm{tr}\, D(h)\ \mathcal{L}(\mathbf{h}^{-1}\tilde{x},\tilde{x}) \tag{17.8}$$

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 fig. 10.4. $\mathrm{tr}\, D(h)\tilde{\mathcal{L}}$ 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

$$(1-t_p)^{m_p} = \det\,(1 - D(h_{\tilde{p}})t_{\tilde{p}})\ , \tag{17.9}$$

valid cycle by cycle in the Euler products (8.12) 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

$$t_p = t_{\tilde{p}}^{h_p} \tag{17.10}$$

computed on the irreducible segment that coresponds to a fundamental domain cycle. For example, in fig. 10.4 we see by inspection that $t_{12} = t_0^2$ and $t_{123} = t_1^3$.

17.3.1 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 stability matrix $\mathbf{J}_p$ of the linearization perpendicular to the orbit and thus the eigenvalues. Typically, *e.g.* if the symmetry is a reflection, some eigenvalues of $\mathbf{J}_p$ change sign. This means that instead of a weight $1/\det(\mathbf{1}-\mathbf{J}_p)$ as for a regular orbit, boundary cycles also pick up contributions of form $1/\det(\mathbf{1}-\mathbf{h}\mathbf{J}_p)$, where $\mathbf{h}$ is a symmetry operation that leaves the orbit pointwise invariant; see for example sect. 17.1.2.

Consequences for the dynamical zeta function factorizations are that sometimes a boundary orbit does not contribute. A derivation of a dynamical zeta function (8.12) from a determinant like (8.9) 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 (8.12). 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 (17.9) will be replaced by a projector. If the weights are insensitive to the signs of the eigenvalues, then this projector is

$$g_p = \frac{1}{h}\sum_{i=1}^{h} b_i \,. \tag{17.11}$$

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 neigborhood. Then (17.11) is replaced by

$$g_p = \frac{1}{h}\sum_{i=1}^{h} \epsilon(b_i)\, b_i \,. \tag{17.12}$$

We have illustrated the above in sect. 17.1.2 by working out the full factorization for the 1-dimensional reflection symmetric maps.

17.4 Factorizations of dynamical zeta functions

In the above 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.

17.4.1 Factorizations of dynamical dynamical zeta functions

According to (17.9) and (17.10), 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 $\tilde{p}$:

$$(1 - t_{\tilde{p}}^{h_p})^{g/h_p} = \det\,(1 - D(h_{\tilde{p}})t_{\tilde{p}}) \tag{17.13}$$

Let $D(h) = \bigoplus_\alpha d_\alpha D_\alpha(h)$ be the decomposition of the matrix representation $D(h)$ into the d_α dimensional irreducible representations α of a finite group G. Such decompositions are block-diagonal, so the corresponding contribution to the Euler product (8.9) factorizes as

$$\det\,(1 - D(h)t) = \prod_\alpha \det\,(1 - D_\alpha(h)t)^{d_\alpha}\ , \tag{17.14}$$

where now the product extends over all distinct d_α-dimensional irreducible representations, each contributing d_α 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 (17.13); 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 (8.12) factorize as

$$\frac{1}{\zeta} = \prod_\alpha \frac{1}{\zeta_\alpha^{d_\alpha}}\ , \qquad \frac{1}{\zeta_\alpha} = \prod_{\tilde{p}} \det\,(1 - D_\alpha(h_{\tilde{p}})t_{\tilde{p}})\ . \tag{17.15}$$

Determinants of d-dimensional irreducible representations can be evaluated using the expansion of determinants in terms of traces,

$$\det\,(1 + M) \;=\; 1 + \mathrm{tr}\,M + \frac{1}{2}\left((\mathrm{tr}\,M)^2 - \mathrm{tr}\,M^2\right)$$

$$+\frac{1}{6}\left((\mathrm{tr}\,M)^3 - 3\,(\mathrm{tr}\,M)(\mathrm{tr}\,M^2) + 2\,\mathrm{tr}\,M^3\right)$$
$$+\cdots+\frac{1}{d!}\left((\mathrm{tr}\,M)^d - \cdots\right)\ , \tag{17.16}$$

(see (J.23), for example) and each factor in (17.14) can be evaluated by looking up the characters $\chi_\alpha(h) = \mathrm{tr}\,D_\alpha(h)$ in standard tables [14]. In terms of characters, we have for the 1-dimensional representations

$$\det\left(1 - D_\alpha(h)t\right) = 1 - \chi_\alpha(h)t\ ,$$

for the 2-dimensional representations

$$\det\left(1 - D_\alpha(h)t\right) = 1 - \chi_\alpha(h)t + \frac{1}{2}\left(\chi_\alpha(h)^2 - \chi_\alpha(h^2)\right)t^2,$$

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.

17.2 on p. 403

17.4.2 Factorizations of spectral determinants

Factorization of the full spectral determinant (8.3) proceeds in essentially the same manner as the factorization of dynamical zeta functions outlined above. By (17.5) and (17.8) 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 (8.9) into

$$\begin{aligned} F(z) &= \prod_\alpha F_\alpha(z)^{d_\alpha} \\ F_\alpha(z) &= \exp\left(-\sum_{\tilde{p}}\sum_{r=1}^{\infty}\frac{1}{r}\frac{\chi_\alpha(h_{\tilde{p}}^r)z^{n_{\tilde{p}}r}}{\left|\det\left(\mathbf{1}-\tilde{\mathbf{J}}_{\tilde{p}}^r\right)\right|}\right)\ , \end{aligned} \tag{17.17}$$

where $\tilde{\mathbf{J}}_{\tilde{p}} = \mathbf{h}_{\tilde{p}}\mathbf{J}_{\tilde{p}}$ is the fundamental domain Jacobian. Boundary orbits require special treatment, discussed in sect. 17.3.1, with examples given in the next section as well as in the specific factorizations discussed below.

The factorizations (17.15), (17.17) 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.

17.5 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 labelled 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.

17.5 on p. 404

The symmetry reduced labelling $\rho_i \in \{0, 1\}$ is related to the standard $s_i \in \{+, -\}$ Ising spin labelling by

$$\begin{array}{lllll} \text{If} & s_i & = & s_{i-1} & \text{then} \quad \rho_i = 1 \\ \text{If} & s_i & \neq & s_{i-1} & \text{then} \quad \rho_i = 0 \end{array} \tag{17.18}$$

For example, $\overline{+} = \cdots+++\cdots$ maps into $\cdots111\cdots = \overline{1}$ (and so does $\overline{-}$), $\overline{-+} = \cdots-+-+\cdots$ maps into $\cdots000\cdots = \overline{0}$, $\overline{-++-} = \cdots--++--++\cdots$ maps into $\cdots0101\cdots = \overline{01}$, and so forth. A list of such reductions is given in table 17.1.

Depending on the maximal symmetry group $\mathcal{H}_p$ that leaves an orbit p invariant (see sects. 17.2 and 17.3 as well as sect. 17.1.2), the contributions to the dynamical zeta function factor as

$$\begin{array}{lllll} & & & A_1 & A_2 \\ \mathcal{H}_p = \{e\}: & (1-t_{\tilde{p}})^2 & = & (1-t_{\tilde{p}})(1-t_{\tilde{p}}) & \\ \mathcal{H}_p = \{e,\sigma\}: & (1-t_{\tilde{p}}^2) & = & (1-t_{\tilde{p}})(1+t_{\tilde{p}}) & , \end{array} \tag{17.19}$$

$\tilde{p}$	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

Table 17.1: 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.

For example:

$$\begin{aligned}
\mathcal{H}_{++-} = \{e\}: \quad (1-t_{++-})^2 &= (1-t_{001})(1-t_{001}) \\
\mathcal{H}_{+-} = \{e,\sigma\}: \quad (1-t_{+-}) &= (1-t_0)\ (1+t_0), \quad t_{+-} = t_0^2
\end{aligned}$$

This yields two binary cycle expansions. The A_1 subspace dynamical zeta function is given by the standard binary expansion (13.5). The antisymmetric A_2 subspace dynamical zeta function ζ_{A_2} differs from ζ_{A_1} only by a minus sign for cycles with an odd number of 0's:

$$\begin{aligned}
1/\zeta_{A_2} &= (1+t_0)(1-t_1)(1+t_{10})(1-t_{100})(1+t_{101})(1+t_{1000}) \\
&\quad (1-t_{1001})(1+t_{1011})(1-t_{10000})(1+t_{10001}) \\
&\quad (1+t_{10010})(1-t_{10011})(1-t_{10101})(1+t_{10111})\dots \\
&= 1+t_0-t_1+(t_{10}-t_1t_0)-(t_{100}-t_{10}t_0)+(t_{101}-t_{10}t_1) \\
&\quad -(t_{1001}-t_1t_{001}-t_{101}t_0+t_{10}t_0t_1)-\dots\dots
\end{aligned} \tag{17.20}$$

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. 17.3.1) with group-theoretic factor $\mathbf{h}_p = (\mathbf{e}+\sigma)/2$, the boundary orbit does not contribute to the antisymmetric subspace

$$\begin{aligned}
&&A_1 \qquad A_2 \\
\text{boundary:} \quad (1-t_p) &= (1-t_{\tilde{p}})(1-0t_{\tilde{p}})
\end{aligned} \tag{17.21}$$

This is the $1/\zeta$ part of the boundary orbit factorization of sect. 17.1.2.

17.6 C_{3v} factorization: 3-disk game of pinball

The next example, the C_{3v} symmetry, can be worked out by a glance at fig. 10.4a. 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 fig. 10.4b). 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 10.2.

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 (17.14) factorizes according to

$$\det\left(1-D(h)t\right) = \left(1-\chi_{A_1}(h)t\right)\left(1-\chi_{A_2}(h)t\right)\left(1-\chi_E(h)t+\chi_{A_2}(h)t^2\right)^2 \tag{17.22}$$

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 [14]

C_{3v}	A_1	A_2	E
e	1	1	2
C_3, C_3^2	1	1	-1
σ_v	1	-1	0

into (17.22), we obtain for the three classes of possible orbit symmetries (indicated in the first column)

$$\begin{array}{rrcll} & \mathbf{h}_{\tilde{p}} & & A_1 \quad A_2 \quad E & \\ e: & (1-t_{\tilde{p}})^6 & = & (1-t_{\tilde{p}})(1-t_{\tilde{p}})(1-2t_{\tilde{p}}+t_{\tilde{p}}^2)^2 & \\ C_3, C_3^2: & (1-t_{\tilde{p}}^3)^2 & = & (1-t_{\tilde{p}})(1-t_{\tilde{p}})(1+\ t_{\tilde{p}}+t_{\tilde{p}}^2)^2 & \\ \sigma_v: & (1-t_{\tilde{p}}^2)^3 & = & (1-t_{\tilde{p}})(1+t_{\tilde{p}})(1+0t_{\tilde{p}}-t_{\tilde{p}}^2)^2. & \end{array} \tag{17.23}$$

where σ_v stands for any one of the three reflections.

The Euler product (8.12) on each irreducible subspace follows from the factorization (17.23). On the symmetric A_1 subspace the ζ_{A_1} is given by the standard binary curvature expansion (13.5). The antisymmetric A_2 subspace ζ_{A_2} differs from ζ_{A_1} only by a minus sign for cycles with an odd number of 0's, and is given in (17.20). For the mixed-symmetry subspace E the curvature expansion is given by

$$\begin{aligned} 1/\zeta_E &= (1+zt_1+z^2t_1^2)(1-z^2t_0^2)(1+z^3t_{100}+z^6t_{100}^2)(1-z^4t_{10}^2) \\ &\quad (1+z^4t_{1001}+z^8t_{1001}^2)(1+z^5t_{10000}+z^{10}t_{10000}^2) \\ &\quad (1+z^5t_{10101}+z^{10}t_{10101}^2)(1-z^5t_{10011})^2\dots \\ &= 1+zt_1+z^2(t_1^2-t_0^2)+z^3(t_{001}-t_1t_0^2) \\ &\quad +z^4\left[t_{0011}+(t_{001}-t_1t_0^2)t_1-t_{01}^2\right] \\ &\quad +z^5\left[t_{00001}+t_{01011}-2t_{00111}+(t_{0011}-t_{01}^2)t_1+(t_1^2-t_0^2)t_{100}\right] \end{aligned} \tag{17.24}$$

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.

Refering back to the topological polynomial (11.30) obtained by setting $t_p = 1$, we see that its factorization is a consequence of the C_{3v} factorization of the ζ function:

$$1/\zeta_{A_1} = 1-2z\ , \quad 1/\zeta_{A_2} = 1\ , \quad 1/\zeta_E = 1+z\ , \tag{17.25}$$

as obtained from (13.5), (17.20) and (17.24) for $t_p = 1$.

Their symmetry is $K = \{\mathbf{e}, \sigma\}$, so according to (17.11), 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 $\det\left(1-\frac{\mathbf{e}+\sigma}{2}t_{\tilde{p}}\right)$ yields

$$\begin{aligned} & & A_1 \quad & A_2 \quad & E & \\ \text{boundary:}\quad (1-t_p)^3 &= (1-t_{\tilde{p}})(1-0t_{\tilde{p}})(1-t_{\tilde{p}})^2\,, & & t_p = t_{\tilde{p}}\,. \end{aligned} \tag{17.26}$$

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;

$$\begin{aligned} & & A_1 \quad & A_2 \quad & E & \\ \text{boundary:}\quad (1-t_p)^3 &= (1-0t_{\tilde{p}})(1-t_{\tilde{p}})(1-t_{\tilde{p}})^2\,, & & t_p = t_{\tilde{p}}\,. \end{aligned} \tag{17.27}$$

Commentary

Remark **17.1** Some examples of systems with discrete symmetries. This chapter is based on ref. [1]. One has a C_2 symmetry in the Lorenz system [1, 15], the Ising model, and in the 3-dimensional anisotropic Kepler potential [24, 38, 39], a C_{3v} symmetry in Hénon-Heiles type potentials [2, 6, 7, 5], a C_{4v} symmetry in quartic oscillators [9, 10], in the pure x^2y^2 potential [11, 12] and in hydrogen in a magnetic field [13], and a $C_{2v} = C_2 \times C_2$ symmetry in the stadium billiard [4]. A very nice application of the symmetry factorization is carried out in ref. [8].

Remark **17.2** Who did it? 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. [9]. They are closely related to the symmetrizations introduced by Gutzwiller [24] in the context of the semiclassical periodic orbit trace formulas, put into more general group-theoretic context by Robbins [4], whose exposition, together with Lauritzen's [5] treatment of the boundary orbits, has influenced the presentation given here. A related group-theoretic decomposition in context of hyperbolic billiards was utilized in ref. [8].

Remark **17.3** Computations The techniques of this chapter have been applied to computations of the 3-disk classical and quantum spectra in refs. [2, **?**], and to a "Zeeman effect" pinball and the x^2y^2 potentials in refs. [3, 11]. 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 [8, 4, **?**] for hyperbolic billiards [22] and Selberg zeta functions [28].

Remark **17.4** 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.

Remark **17.5** Cycles and symmetries. We conclude this section with a few comments about the role of symmetries in actual extraction of cycles. In the example at hand, the N-disk billiard systems, a fundamental domain is a sliver of the N-disk configuration space delineated by a pair of adjoining symmetry axes, with the directions of the momenta indicated by arrows. The flow may further be reduced to a return map on a Poincaré surface

of section, on which an appropriate transfer operator may be constructed. While in principle any Poincaré surface of section will do, a natural choice in the present context are crossings of symmetry axes.

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.

Remark **17.6** C_2 symmetry The C_2 symmetry arises, for example, in the Lorenz system [15], in the 3-dimensional anisotropic Kepler problem [24, 38, 39] or in the cycle expansions treatments of the Ising model [**?**].

Remark **17.7** Hénon-Heiles potential An example of a system with C_{3v} symmetry is provided by the motion of a particle in the Hénon-Heiles potential [2]

$$V(r,\theta) = \frac{1}{2}r^2 + \frac{1}{3}r^3\sin(3\theta)\ .$$

Our coding is not directly applicable to this system because of the existence of elliptic islands and because the three orbits that run along the symmetry axis cannot be labelled in our code. However, since these orbits run along the boundary of the fundamental domain, they require the special treatment discussed in sect. 17.3.1.

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.

Reduction to the fundamental domain simplifies symbolic dynamics and eliminates symmetry induced degeneracies. While the resummation of the theory 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, accomodating 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 phase 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 phase space, *i.e.*, an increase in density by a factor 3 with the same numerical effort.

We emphasize that the symmetry factorization (17.23) of the dynamical zeta functionis *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 [4]. In conclusion, the manifold advantages of the symmetry reduced dynamics should thus be obvious; full space cycle expansions, such as those of exercise 13.8, are useful only for cross checking purposes.

References

[17.1] P. Cvitanović and B. Eckhardt, *"Symmetry decomposition of chaotic dynamics"*, *Nonlinearity* **6**, 277 (1993).

[17.2] M. Henón and C. Heiles, *J. Astron.* **69**, 73 (1964).

[17.3] G. Russberg, (in preparation)

[17.4] J.M. Robbins, *Phys. Rev.* **A 40**, 2128 (1989).

[17.5] B. Lauritzen, Discrete symmetries and the periodic-orbit expansions, *Phys. Rev.* **A 43** 603, (1991).

[17.6] C. Jung and H.J. Scholz, *J. Phys.* **A 20**, 3607 (1987).

[17.7] C. Jung and P. Richter, *J. Phys.* **A 23**, 2847 (1990).

[17.8] N. Balasz and A. Voros, *Phys. Rep.* **143**, 109 (1986).

[17.9] B. Eckhardt, G. Hose and E. Pollak, *Phys. Rev.* **A 39**, 3776 (1989).

[17.10] C. C. Martens, R. L. Waterland, and W. P. Reinhardt, *J. Chem. Phys.* **90**, 2328 (1989).

[17.11] S.G. Matanyan, G.K. Savvidy, and N.G. Ter-Arutyunyan-Savvidy, *Sov. Phys. JETP 53*, 421 (1981).

[17.12] A. Carnegie and I. C. Percival, *J. Phys.* **A 17**, 801 (1984).

[17.13] B. Eckhardt and D. Wintgen, *J. Phys.* **B 23**, 355 (1990).

[17.14] M. Hamermesh, *Group Theory and its Application to Physical Problems* (Addison-Wesley, Reading, 1962).

[17.15] G. Ott and G. Eilenberger, private communication.

Exercises

17.1 Sawtooth map desymmetrization. Work out the some of the shortest global cycles of different symmetries and fundamental domain cycles for the sawtooth map of fig. 17.1. Compute the dynamical zeta function and the spectral determinant of the Perron-Frobenius operator for this map; check explicitely the factorization (17.2).

17.2 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 (17.16) is the trace of the antisymmetric part of the $M \times M$ Kronecker product; if α is a 2-dimensional representation, this is the A_2 antisymmetric representation, so

$$\text{2-dim:} \quad \det\left(1 - D_\alpha(h)t\right) = 1 - \chi_\alpha(h)t + \chi_{A_2}(h)t^2. \tag{17.28}$$

17.3 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).

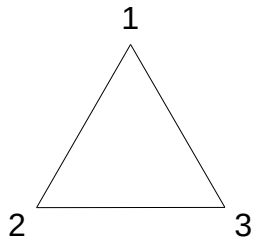
b) Find the shortest cycle with no symmetries and factorize it like in a)

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 (17.15) and (17.17).

(Per Rosenqvist)

17.4 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) 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) Their are three irreducible representations for the group. Two are one dimensional and the other one is formed by 2×2 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.

17.5 C_2 factorizations: the Lorenz and Ising systems. In the Lorenz system [1, **?**, 15] the labels + and − stand for the left or the right lobe of the attractor and the symmetry is a rotation by π around the z-axis. Similarly, the Ising Hamiltonian (in the absence of an external magnetic field) is invariant under spin flip. Work out the factorizations for some of the short cycles in either system.

17.6 Ising model. The Ising model with two states $\epsilon_i = \{+, -\}$ per site, periodic boundary condition, and Hamiltonian

$$H(\epsilon) = -J \sum_i \delta_{\epsilon_i, \epsilon_{i+1}} ,$$

is invariant under spin-flip: $+ \leftrightarrow -$. Take advantage of that symmetry and factorize the dynamical zeta function for the model, that is, find all the periodic orbits that contribute to each factor and their weights.

17.7 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.

Chapter 18

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, especially for systems near or far from equilibrium.

Sinai, Ruelle and Bowen (SRB) have generalized Boltzmann's notion of ergodicity for a constant energy surface for a Hamiltonian system in equilibrium to a dissipative system in a nonequilibrium stationary state. In this more general setting the attractor plays the role of a constant energy surface, and the SRB measure of sect. 5.1 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 the diffusion constant and the pressure, to which we turn now. We shall apply cycle expansions to the analysis of *transport* properties of chaotic systems. The infinite extent systems for which the periodic orbit theory yields formulas for diffusion and other transport coefficients are spatially periodic, the global phase 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 two dimensional rotating flows, problems which can be described as deterministic diffusion in periodic arrays.

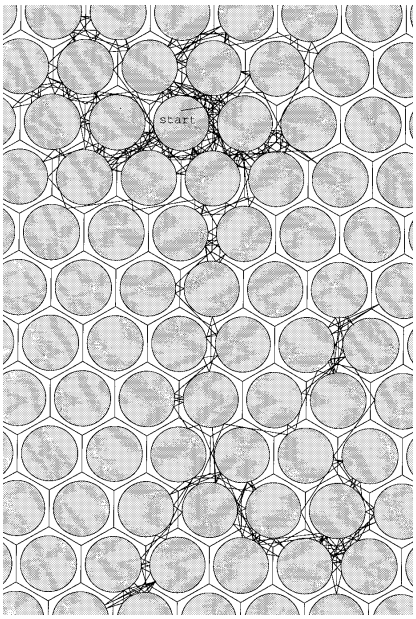


Figure 18.1: Deterministic diffusion in a finite horizon periodic Lorentz gas. (Courtesy of T. Schreiber)

18.1 Diffusion in periodic arrays

The 2-dimensional *Lorentz gas* is an infinite scatterer array in which diffusion of a light molecule in a gas of heavy scatterers is modelled 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 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, fig. 18.1. 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 phase 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 phase 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 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 fig. 18.2. We distinguish two types

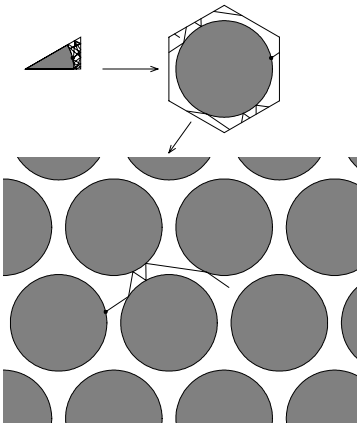


Figure 18.2: Tiling of $\hat{\mathcal{M}}$, a periodic lattice of reflecting disks, by the fundamental domain $\widetilde{\mathcal{M}}$. Indicated is an example of a global trajectory $\hat{x}(t)$ together with the corresponding elementary cell trajectory $x(t)$ and the fundamental domain trajectory $\tilde{x}(t)$. (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. **??**.

As we will work with three kinds of phase spaces, good manners require that we repeat what hats, tildas and nothings atop symbols signify:

$$\begin{array}{ll} \tilde{} & \text{fundamental domain, triangle in fig. 18.2} \\ & \text{elementary cell, hexagon in fig. 18.2} \\ \hat{} & \text{full phase space, lattice in fig. 18.2} \end{array} \tag{18.1}$$

It is convenient to define an evolution operator for each of the 3 cases of fig. 18.2. $\hat{x}(t) = \hat{f}^t(\hat{x})$ 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

$$\hat{n}_t(x_0) = \hat{f}^t(x_0) - f^t(x_0) \in T\,, \tag{18.2}$$

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 $x(t) \in \mathcal{M}$ (see chapter 17).

Fix a vector $\beta \in \mathbb{R}^d$, where d is the dimension of the phase space. We will compute the diffusive properties of the Lorentz gas from the expectation value

(6.11)

$$s(\beta) \;=\; \lim_{t\to\infty} \frac{1}{t} \log\langle e^{\beta\cdot(\hat{x}(t)-x)}\rangle_{\mathcal{M}} \;, \tag{18.3}$$

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

$$\frac{\partial}{\partial\beta_i}\frac{\partial}{\partial\beta_j} s(\beta)\Big|_{\beta=0} \;=\; \lim_{t\to\infty} \frac{1}{t}\langle(\hat{x}(t)-x)_i(\hat{x}(t)-x)_j\rangle_{\mathcal{M}} \;,$$

yield a (generally anisotropic) diffusion matrix. The spatial diffusion constant is then given by the Einstein relation

$$D \;=\; \frac{1}{2d}\sum_i \frac{\partial^2}{\partial\beta_i^2} s(\beta)\Big|_{\beta=0} \;=\; \lim_{t\to\infty} \frac{1}{2dt}\langle(\hat{q}(t)-q)^2\rangle_{\mathcal{M}} \quad ,$$

where the i sum is restricted to the spatial components q_i of the phase space vectors $x = (q, p)$.

We now turn to the connection between (18.3) 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 (see one of the final remarks).

18.1.1 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. 6.2, 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 $y \in \mathcal{M}$. Therefore, and this is the main point, translation invariance can be used to reduce this average to the elementary cell:

$$\langle e^{\beta\cdot(\hat{x}(t)-x)}\rangle_{\mathcal{M}} \;=\; \frac{1}{|\mathcal{M}|}\int_{x,y\in\mathcal{M}} dx dy\, e^{\beta\cdot(\hat{f}^t(x)-x)}\delta(y-f^t(x)) \;. \tag{18.4}$$

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 (18.4) suggests that we study the evolution operator

$$\mathcal{L}^t(y,x) \;=\; e^{\beta\cdot(\hat{x}(t)-x)}\delta(y-f^t(x))\;, \tag{18.5}$$

where $\hat{x}(t) \;=\; \hat{f}^t(x) \;\in\; \hat{\mathcal{M}}$, but $x, x(t), y \in \mathcal{M}$. It is straightforward to check that this operator has the semigroup property (6.21), $\int_{\mathcal{M}} dz\, \mathcal{L}^{t_2}(y,z)\mathcal{L}^{t_1}(z,x) = \mathcal{L}^{t_2+t_1}(y,x)$. For $\beta = 0$, the operator (18.5) is the Perron-Frobenius operator (5.10), with the leading eigenvalue $e^{s_0} = 1$ by the flow conservation sum rule (14.11).

The rest is old hat. As in sect. 7.1.4, the spectrum of $\mathcal{L}$ is evaluated by taking the trace

$$\mathrm{tr}\,\mathcal{L}^t \;=\; \int_{\mathcal{M}} dx\, e^{\beta\cdot\hat{n}_t(x)}\delta(x-x(t))\;.$$

Here $\hat{n}_t(x)$ is the discrete lattice translation defined in (18.2). 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 phase 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 phase 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 (8.6)

$$F(\beta,s) \;=\; \prod_p \exp\left(-\sum_{r=1}^{\infty}\frac{1}{r}\frac{e^{(\beta\cdot\hat{n}_p-sT_p)r}}{\left|\det\left(\mathbf{1}-\mathbf{J}_p^r\right)\right|}\right)\;, \tag{18.6}$$

or the corresponding dynamical zeta function (8.12)

$$1/\zeta(\beta,s) \;=\; \prod_p\left(1-\frac{e^{(\beta\cdot\hat{n}_p-sT_p)}}{|\Lambda_p|}\right)\;. \tag{18.7}$$

The associated dynamical zeta function cycle averaging formula (13.17) for the diffusion constant (6.13), zero mean drift $\langle\hat{x}_i\rangle = 0$, is given by

$$D \;=\; \frac{1}{2d}\frac{\langle\hat{x}^2\rangle_\zeta}{\langle T\rangle_\zeta} \;=\; \frac{1}{2d}\frac{1}{\langle T\rangle_\zeta}\frac{\sum'(-1)^{k+1}(\hat{n}_{p_1}+\cdots+\hat{n}_{p_k})^2}{|\Lambda_{p_1}\cdots\Lambda_{p_k}|}\,. \tag{18.8}$$

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 accross an infinite lattice; nevertheless, the reduction to the elementary cell enables us to compute relevant quantities in the usual way, in terms of periodic orbits. A sleepy reader might protest that $x_p = x(T_p) - x(0)$ 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 $\langle \hat{x}^2 \rangle_\zeta$ gets contributions only from the periodic runaway trajectories; they are closed in the elementary cell, but on the periodic lattice each one grows like $\hat{x}(t)^2 = (t/T_p)^2 \hat{n}_p^2 \sim t^2$. 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 $1/|\Lambda_p| \approx e^{-T_p \lambda}$ - here λ is the Lyapunov exponent-), and they may lead to anomalous diffusion (accelerated or slowed down depending whether running or standing orbits are characterized by enhanced probabilities), see sect. **??**.

To illustrate the main idea, tracking of a globally diffusing orbit by the associated confined orbit restricted to the elementary cell, we start with a class of simple 1-d dynamical systems where all transport coefficients can be evaluated analytically. If you would like to master the material, working through the project L.1 and or project L.2 is strongly recommended. We return to the Lorentz gas in sect. **??**.

18.2 Diffusion induced by chains of 1-d maps

In a typical deterministic diffusive process, trajectories originating from a given scatterer reach some number of neighboring scatterers in one bounce, and then the process is repeated. As was shown in chapter **??**, 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 modelled by 1-dimensional expanding maps. This observation motivates introduction of a class of particularly simple 1-dimensional systems, chains of piecewise linear maps.

We start by defining the map $\hat{f}$ on the unit interval as

$$\hat{f}(\hat{x}) = \begin{cases} \Lambda \hat{x} & \hat{x} \in [0, 1/2) \\ \Lambda \hat{x} + 1 - \Lambda & \hat{x} \in (1/2, 1] \end{cases} , \qquad \Lambda > 2 \,, \tag{18.9}$$

and then extending the dynamics to the entire real line, by imposing the translation property

$$\hat{f}(\hat{x}+\hat{n}) \;=\; \hat{f}(\hat{x})+\hat{n} \qquad \hat{n}\in\mathbb{Z}\,. \tag{18.10}$$

As the map is dicontinuous 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. Even though this means omitting a single point, the consequences for the symbolic dynamics can be significant, as will be evident in the derivation of the diffusion constant formula (18.20).

The map is symmetric under the $\hat{x}$-coordinate flip

$$\hat{f}(\hat{x}) \;=\; -\hat{f}(-\hat{x})\,, \tag{18.11}$$

so the dynamics will exhibit no mean drift; all odd derivatives (with respect to β) of the generating function (6.11) evaluated at $\beta=0$ will vanish.

The map (18.9) is sketched in fig. 18.3(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 Λ, other points jump $\hat{n}$ cells, either to the right or to the left. Repetition of this process generates a trajectory that for long times is essentially a random walk for almost every initial condition.

The translational symmetry (18.10) 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

$$f(x)=\hat{f}(\hat{x})-\left[\hat{f}(\hat{x})\right]\,, \qquad x=\hat{x}-[\hat{x}]\in[0,1] \tag{18.12}$$

fig. 18.3(b), where $[\cdots]$ stands for the integer part. We showed in the former section that elementary cell cycles either correspond to 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 $t_p \;=\; z^{n_p}e^{\beta\hat{n}_p}/|\Lambda_p|$.

For the piecewise linear map of fig. 18.3 we can evaluate the dynamical zeta function in closed form. Each branch has the same value of the slope, and the map can be parametrized either by its critical value $a=\hat{f}(1/2)$, the absolute maximum on the interval $[0,1]$ related to the slope of the map by $a=\Lambda/2$, or by the slope Λ, the stretching of the map.

The diffusion constant formula (18.8) for 1-d maps is

$$D=\frac{1}{2}\frac{\langle\hat{n}^2\rangle_\zeta}{\langle n\rangle_\zeta} \tag{18.13}$$

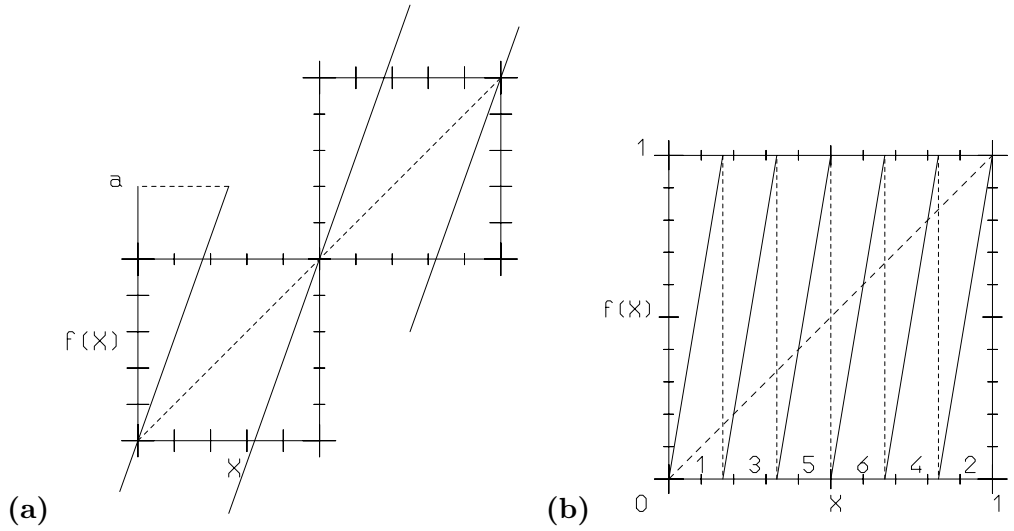


Figure 18.3: (a) $\hat{f}(\hat{x})$, the full sawtooth map (18.9). (b) $f(x)$, the sawtooth map restricted to the unit circle (18.12), $\Lambda = 6$.

where the "mean cycle time" is given by

$$\langle n \rangle_\zeta = z \frac{\partial}{\partial z} \frac{1}{\zeta(0,z)} \bigg|_{z=1} = -\sum{}' (-1)^k \frac{n_{p_1} + \cdots + n_{p_k}}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|} \,, \tag{18.14}$$

the mean cycle displacement squared by

$$\langle \hat{n}^2 \rangle_\zeta = \frac{\partial^2}{\partial \beta^2} \frac{1}{\zeta(\beta,1)} \bigg|_{\beta=0} = -\sum{}' (-1)^k \frac{(\hat{n}_{p_1} + \cdots + \hat{n}_{p_k})^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|} \,, \tag{18.15}$$

the sum being again over all distinct non-repeating combinations of prime cycles. The evaluation of these formulas in what follows in this section will require no more than pencil and paper computations.

18.2.1 Case of unrestricted symbolic dynamics

Whenever Λ is an integer number, the symbolic dynamics can be easily characterized. For example, for the case $\Lambda = 6$ illustrated in fig. 18.3(b), the circle map consists of 6 full branches, with uniform stretching factor $\Lambda = 6$. The branches have different jumping numbers: for branches 1 and 2 we have $\hat{n} = 0$, for branch 3 we have $\hat{n} = +1$, for branch 4 $\hat{n} = -1$, and finally for branches 5 and 6 we have respectively $\hat{n} = +2$ and $\hat{n} = -2$. The same structure reappears whenever Λ is an even integer $\Lambda = 2a$: all branches are mapped onto the whole unit interval and we

have two $\hat{n} = 0$ branches, one branch for which $\hat{n} = +1$ and one for which $\hat{n} = -1$, and so on, up to the maximal jump $|\hat{n}| = a - 1$. The symbolic dynamics is thus full, unrestricted shift in $2a$ letters $\{0_+, 1_+, \ldots, (a-1)_+, (a-1)_-, \ldots, 1_-, 0_-\}$, where the letter indicates both the length and the direction of the corresponding jump.

For the piecewise linear maps with uniform stretching the weight of a symbol sequence is a product of weights for individual steps, $t_{sq} = t_s t_q$. For the map of fig. 18.3 there are 6 distinct weigths:

$$\begin{aligned} t_1 &= t_2 = z/\Lambda \\ t_3 &= e^{\beta} z/\Lambda\,, \quad t_4 = e^{-\beta} z/\Lambda\,, \quad t_5 = e^{2\beta} z/\Lambda\,, \quad t_6 = e^{-2\beta} z/\Lambda\,. \end{aligned}$$

We now take full advantage of the piecewise linearity and of the simple structure of the symbolic dynamics, that lead to full cancellation of all curvature corrections in (13.5), and write down the *exact* dynamical zeta function (11.12) just in terms of the fixed point contributions:

$$\begin{aligned} 1/\zeta(\beta, z) &= 1 - t_{0_+} - t_{0_-} - \cdots - t_{(a-1)_+} - t_{(a-1)_-} \\ &= 1 - \frac{z}{a}\left(1 + \sum_{j=1}^{a-1} \cosh(\beta j)\right)\,. \end{aligned} \tag{18.16}$$

The leading (and only) eigenvalue of the evolution operator (18.5) is

$$s(\beta) = \log\left\{\frac{1}{a}\left(1 + \sum_{j=1}^{a-1} \cosh(\beta j)\right)\right\}\,, \qquad \Lambda = 2a \text{ even integer}\,. \tag{18.17}$$

Evidently, as required by the flow conservation (14.11), $s(0) = 0$. 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 (18.13):

$$\langle T \rangle_\zeta = 2a\frac{1}{\Lambda} = 1\,, \qquad \left\langle \hat{x}^2 \right\rangle_\zeta = 2\frac{0^2}{\Lambda} + 2\frac{1^2}{\Lambda} + 2\frac{2^2}{\Lambda} + \cdots + 2\frac{(a-1)^2}{\Lambda} \tag{18.18}$$

Using the identity $\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$ we obtain

$$D = \frac{1}{24}(\Lambda - 1)(\Lambda - 2)\,, \qquad \Lambda \text{ even integer}\,. \tag{18.19}$$

Similar calculation for odd integer $\Lambda = 2k-1$ yields 18.1 on p. 424

$$D = \frac{1}{24}(\Lambda^2 - 1)\,, \qquad \Lambda \text{ odd integer}\,. \tag{18.20}$$

18.2.2 Higher order transport coefficients

The same approach yields higher order transport coefficients

$$\mathcal{B}_k = \frac{1}{k!}\frac{d^k}{d\beta^k}s(\beta)\Big|_{\beta=0}\,, \qquad \mathcal{B}_2 = D\,. \tag{18.21}$$

The behavior of the higher order coefficients yields information on the form of the asymptotic distribution function generated by the diffusive process. We remark that here $\hat{x}_t$ is the relevant dynamical variable (and not the time integral of the observable we are interested in like in (6.1)), so the generating function actually provides information about moments of arbitrary orders. Were the diffusive process purely gaussian

$$e^{ts(\beta)} = \frac{1}{\sqrt{4\pi Dt}}\int_{-\infty}^{+\infty} d\sigma e^{\beta\sigma}e^{-\sigma^2/(4Dt)} = e^{\beta^2 Dt} \tag{18.22}$$

the only 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 $\mathcal{B}_4$, a measure of deviation from gaussian behavior, is easily evaluated. Using (18.17) in the case of even integer slope $\Lambda = 2a$ we obtain 18.2 on p. 424

$$\mathcal{B}_4 = -\frac{1}{4!\cdot 60}(a-1)(2a-1)(4a^2-9a+7)\,. \tag{18.23}$$

Higher order even coefficients may be calculated along the same lines.

18.2.3 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.

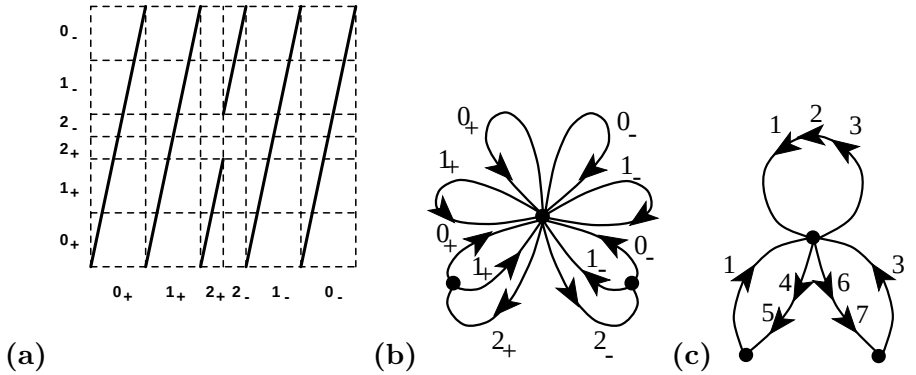


Figure 18.4: (a) A partition of the unit interval into six intervals, labeled by the jumping number $\hat{n}(x)$ $I = \{0_+, 1_+, 2_+, 2_-, 1_-, 0_-\}$. The partition is Markov, as the critical point is mapped onto the right border of $\mathcal{M}_{1_+}$. (b) The Markov graph for this partition. (c) The Markov graph in the compact notation of (18.25) (introduced by Vadim Moroz).

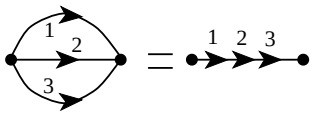
The key idea is to construct a *Markov partition* (10.4) where intervals are mapped *onto* unions of intervals. As an example we determine a value of the parameter $4 \leq \Lambda \leq 6$ for which $f(f(1/2)) = 0$. As in the integer Λ case, we partition the unit interval into six intervals, labeled by the jumping number $\hat{n}(x) \in \{\mathcal{M}_{0_+}, \mathcal{M}_{1_+}, \mathcal{M}_{2_+}, \mathcal{M}_{2_-}, \mathcal{M}_{1_-}, \mathcal{M}_{0_-}\}$, ordered by their placement along the unit interval, fig. 18.4(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 $\mathcal{M}_{2_+}$. Equating $f(1/2)$ with the right border of $\mathcal{M}_{2_+}$, $x = 1/\Lambda$, we obtain a quadratic equation with the expanding solution $\Lambda = 2(\sqrt{2}+1)$. For this parameter value $f(\mathcal{M}_{2_+}) = \mathcal{M}_{0_+} \bigcup \mathcal{M}_{1_+}$, $f(\mathcal{M}_{2_-}) = \mathcal{M}_{0_-} \bigcup \mathcal{M}_{1_-}$, while the remaining intervals map onto the whole unit interval $\mathcal{M}$. The transition matrix (10.2) is given by

$$\phi' = T\phi \;=\; \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \phi_{0_+} \\ \phi_{1_+} \\ \phi_{2_+} \\ \phi_{2_-} \\ \phi_{1_-} \\ \phi_{0_-} \end{pmatrix} \tag{18.24}$$

One could diagonalize (18.24) on a computer, but, as we saw in sect. 10.8, the Markov graph fig. 18.4(b) corresponding to fig. 18.4(a) offers more insight into the dynamics. The graph fig. 18.4(b) can be redrawn more compactly as Markov

graph fig. 18.4(c) by replacing parallel lines in a graph by their sum



$$= t_1 + t_2 + t_3 \,. \tag{18.25}$$

The dynamics is unrestricted in the alphabet

$$\mathcal{A} = \{0_+,\, 1_+,\, 2_+0_+,\, 2_+1_+,\, 2_-1_-,\, 2_-0_-,\, 1_-,\, 0_-\}\,,$$

and we are led to the dynamical zeta function

$$\begin{aligned} 1/\zeta(\beta,z) &= 1 - t_{0_+} - t_{1_+} - t_{2_+0_+} - t_{2_+1_+} - t_{2_-1_-} - t_{2_-0_-} - t_{1_-} - t_{0_-} \\ &= 1 - \frac{2z}{\Lambda}\,(1+\cosh(\beta)) - \frac{2z^2}{\Lambda^2}\,(\cosh(2\beta)+\cosh(3\beta))\ . \end{aligned} \tag{18.26}$$

(see the follows loop expansion (11.12) of sect. 11.3). 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 (14.11). Conversely, we could have started by picking the desired Markov partition, writing down the corresponding dynamical zeta function, and then fixing Λ by the $1/\zeta(0,1) = 0$ condition. For more complicated Markov graphs this approach, together with the factorization (18.28) is very helpful in reducing the order of the polynomial condition that fixes Λ.

18.3 on p. 424

The diffusion constant follows from (18.13)

$$\begin{aligned} \langle n\rangle_\zeta &= 4\frac{1}{\Lambda} + 4\frac{2}{\Lambda^2}\,, \quad \langle \hat{n}^2\rangle_\zeta \,=\, 2\frac{1^2}{\Lambda} + 2\frac{2^2}{\Lambda^2} + 2\frac{3^2}{\Lambda^2} \\ D &= \frac{15+2\sqrt{2}}{16+8\sqrt{2}}\,. \end{aligned} \tag{18.27}$$

It is by now clear how to build an infinite hierarchy of Markov cases, by tuning the slope in such a way that the discontinuity point in the centre is mapped into the fixed point at the origin in a finite number p of steps. By taking higher and higher values for the number p of iterates it is possible to see that Markov parameters are dense, organized in a hierarchy that resembles the way in which

rationals are 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 (18.9) for a random "generic" value of the parameter Λ, for example $\Lambda = 4.5$. The idea is to bracket this value of Λ by the nearby ones, for which higher and higher iterates of the critical value $a = \Lambda/2$ 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 $\Lambda = 4.5$. Some details of how this is accomplished are given in appendix **??** for a related problem, the pruned Bernulli shift. Judging how difficult such problem is already for a tent map (see sect. 11.6 and appendix E.1), this is not likely to take only a week of work.

Expressions like (18.19) may lead to an expectation that the diffusion coefficient (and thus transport properties) are smooth functions of parameters controling the chaoticity of the system. For example, one could hope that the diffusion coefficient increases smoothly and monotonically as the slope Λ of the map (18.9) is increased, or, perhaps more physically, that the diffusion coefficient is a smooth function of the Lyapunov exponent λ. This turns out not to be true: D as a function of Λ is a fractal, nowhere differentiable curve. The dependence of D on the map parameter Λ is rather unexpected - even though for larger Λ more points are mapped outside the unit cell in one iteration, the diffusion constant does not necessarily grow.

This is a consequence of the lack of structural stability of not only generic systems, but also purely hyperbolic systems such as the Lozi map and the 1-d diffusion map (18.9). 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 ininitely many periodic orbits. As far as diffusion is concerned this means that even though local expansion rate is a smooth function of Λ, the number of ways in which the trajectory can re-enter the the initial cell is an irregular function of Λ.

The lesson is that lack of structural stabily implies lack of spectral stability, and no global observable is expected to depend smoothly on the system papramеters.

Commentary

Remark **18.1** Lorentz gas. The original pinball model proposed by Lorentz [3] consisted of randomly, rather than regularly placed scatterers.

Remark **18.2** Who's dun it? Cycle expansions for the diffusion con-

Figure 18.5: The dependence of D on the map parameter Λ is continuous, but not monotone. (From ref. [1]).

stant of a particle moving in a periodic array seem to have been introduced independently by R. Artuso [4] (exact dynamical zeta function for 1-d chains of maps (18.8)), by W.N. Vance [5] (the trace formula (**??**) for the Lorentz gas), and by P. Cvitanović, J.-P. Eckmann, and P. Gaspard [6] (the dynamical zeta function cycle expansion (18.8) applied to the Lorentz gas).

Remark **18.3** Structural stability for D Expressions like (18.19) may lead to an expectation that the diffusion coefficient (and thus transport properties) are smooth functions of the chaoticity of the system (parametrized, for example, by the Lyapunov exponent $\lambda = \ln \Lambda$). This turns out not to be true: D as a function of Λ is a fractal, nowhere differentiable curve. The dependence of D on the map parameter Λ is rather unexpected - even though for larger Λ more points are mapped outside the unit cell in one iteration, the diffusion constant does not necessarily grow. The fractal dependence of diffusion constant on the map parameter is discussed in ref. [7]. 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 [8].

Remark **18.4** Diffusion induced by 1-dimensional maps. We refer the reader to refs. [9, 10] for early work on the deterministic diffusion induced by 1-dimenional maps. The sawtooth map (18.9) was introduced by Grossmann and Fujisaka [11] who derived the integer slope formulas (18.19) for the diffusion constant. The sawtooth map is also discussed in refs. [12].

Remark **18.5** Symmetry factorization in one dimension. In the $\beta = 0$ limit the dynamics (18.11) 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. 17.1.2:

$$\frac{1}{\zeta(0,z)} = \frac{1}{\zeta_s(0,z)}\,\frac{1}{\zeta_a(0,z)}\,, \qquad \frac{\partial}{\partial z}\frac{1}{\zeta} = \frac{1}{\zeta_s}\frac{\partial}{\partial z}\frac{1}{\zeta_a} + \frac{1}{\zeta_a}\frac{\partial}{\partial z}\frac{1}{\zeta_s}\,. \tag{18.28}$$

length	# cycles	$\zeta(0,0)$	λ
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 18.1: Fundamental domain, w=0.3 .

The leading (material flow conserving) eigenvalue $z = 1$ belongs to the symmetric subspace $1/\zeta_s(0,1) = 0$, so the derivatives (18.14) also depend only on the symmetric subspace:

$$\langle n \rangle_\zeta = z\frac{\partial}{\partial z}\frac{1}{\zeta(0,z)}\bigg|_{z=1} = \frac{1}{\zeta_a(0,z)}\, z\frac{\partial}{\partial z}\,\frac{1}{\zeta_s(0,z)}\bigg|_{z=1} \tag{18.29}$$

. Implementing the symmetry factorization is convenient, but not essential, at this computational level.

Remark **18.6** Lorentz gas in the fundamental domain. The vector valued nature of the generating function (18.3) 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. [13], numerical estimates for scalar quantities are reported in table 18.1, taken from ref. [14].

In order to perform the full reduction for diffusion one should express the dynamical zeta function (18.7) in terms of the prime cycles of the fundamental domain $\tilde{\mathcal{M}}$ of the lattice (see fig. 18.2) rather than those of the elementary (Wigner-Seitz) cell $\mathcal{M}$. This problem is complicated by the breaking of the rotational symmetry by the auxilliary vector β, or, in other words, the noncommutativity of translations and rotations: see ref. [6] for a discussion of the problem.

Remark **18.7** Anomalous diffusion. Anomalous diffusion for one dimensional 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.

Résumé

The classical Boltzmann equation for evolution of 1-particle density is based on *stosszahlansatz*, neglect of particle correlations prior to, or after a 2-particle

collision. It is a very good approximate description of dilute gas dynamics, but a difficult starting point for inclusion of systematic corrections. In the theory developed here, no correlations are neglected - they are all included in the cycle averaging formula such as the cycle expansion for the diffusion constant

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

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.

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 developped above in physical settings. In cases of severe pruning the trace formulas and ergodic sampling of dominant cycles might be preferable to the cycle expansions of dynamical zeta functions and systematic enumeration of all cycles.

References

[18.1] J. Machta and R. Zwanzig, *Phys. Rev. Lett.* **50**, 1959 (1983).

[18.2] G.P. Morriss and L. Rondoni, *J. Stat. Phys.* **75**, 553 (1994).

[18.3] H.A. Lorentz, *Proc. Amst. Acad.* **7**, 438 (1905).

[18.4] R. Artuso, *Phys. Lett.* **A 160**, 528 (1991).

[18.5] W.N. Vance, *Phys. Rev. Lett.* **96**, 1356 (1992).

[18.6] P. Cvitanović, J.-P. Eckmann, and P. Gaspard, *Chaos, Solitons and Fractals* **6**, 113 (1995).

[18.7] R. Klages and J.R. Dorfman, *Phys. Rev. Lett.* **74**, 387-390 (1995);

[18.8] J. Lloyd, M. Niemeyer, L. Rondoni and G.P. Morriss, *CHAOS* **5**, 536 (1995).

[18.9] T. Geisel and J. Nierwetberg, *Phys. Rev. Lett.* **48**, 7 (1982).

[18.10] M. Schell, S. Fraser and R. Kapral, *Phys. Rev.* **A 26**, 504 (1982).

[18.11] S. Grossmann, H. Fujisaka, Phys. Rev. **A 26**, 1179 (1982); H. Fujisaka and S. Grossmann, Z. Phys. **B 48**, 261 (1982).

[18.12] P. Gaspard and F. Baras, in M. Mareschal and B.L. Holian, eds., *Microscopic simulations of Complex Hydrodynamic Phenomena* (Plenum, NY 1992).

[18.13] F. Christiansen, Master's Thesis, Univ. of Copenhagen (June 1989)

[18.14] P. Cvitanović, P. Gaspard, and T. Schreiber, "Investigation of the Lorentz Gas in terms of periodic orbits", *CHAOS* **2**, 85 (1992).

[18.15] S. Grossmann and S. Thomae, *Phys. Lett.***A 97**, 263 (1983).

[18.16] R. Artuso, G. Casati and R. Lombardi, *Physica* **A 205**, 412 (1994).

Exercises

18.1 Diffusion for odd integer Λ. Show that when the slope $\Lambda = 2k-1$ in (18.9) is an odd integer, the diffusion constant is given by $D = (\Lambda^2-1)/24$, as stated in (18.20).

18.2 Fourth-order transport coefficient. Verify (18.23). You will need the identity

$$\sum_{k=1}^{n} k^4 = \frac{1}{30} n(n+1)(2n+1)(3n^2+3n-1)\,.$$

18.3 Finite Markov partitions. Verify (18.27).

18.4 Maps with variable peak shape:
Consider the following piecewise linear map

$$f_\delta(x) = \begin{cases} \frac{3x}{1-\delta} & \text{for } x \in \left[0, \frac{1}{3}(1-\delta)\right] \\ \frac{3}{2} - \left(\frac{2}{\delta}\left|\frac{4-\delta}{12} - x\right|\right) & \text{for } x \in \left[\frac{1}{3}(1-\delta), \frac{1}{6}(2+\delta)\right] \\ 1 - \frac{3}{1-\delta}\left(x - \frac{1}{6}(2+\delta)\right) & \text{for } x \in \left[\frac{1}{6}(2+\delta), \frac{1}{2}\right] \end{cases} \tag{18.30}$$

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. [15, 16] for further details.

18.5 Two symbol cycles for the Lorentz gas. Write down the full groups of cycles labelled by two symbols, whose representative elements are (0 6), (1 7), (1 5) and (0 5) respectively.

Appendix L contains several project-length deterministic diffusion exercises.

Chapter 19

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 likelyhood 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 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.

19.1 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*

$$x_{n+1} = f(x_n) = x_n + \Omega - \frac{k}{2\pi}\sin(2\pi x_n) \qquad \mod 1 \ . \tag{19.1}$$

$f(x)$ is assumed to be continuous, have a continuous first derivative, and a continuous second derivative at the inflection point (where the second derivative vanishes). For the generic, physically relevant case (the only one considered here) the inflection is cubic. Here k parametrizes the strength of the nonlinear interaction, and Ω is the *bare* frequency.

The phase space of this map, the unit interval, can be thought of as the elementary cell of the map

$$\hat{x}_{n+1} = \hat{f}(\hat{x}_n) = \hat{x}_n + \Omega - \frac{k}{2\pi}\sin(2\pi\hat{x}_n) \ . \tag{19.2}$$

where $\hat{}$ is used in the same sense as in chapter 18.

The winding number is defined as

$$W(k,\Omega) = \lim_{n\to\infty} (\hat{x}_n - \hat{x}_0)/n. \tag{19.3}$$

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 fig. 19.1

$$x_{n+1} = x_n + \Omega \qquad \mod 1 \ , \tag{19.4}$$

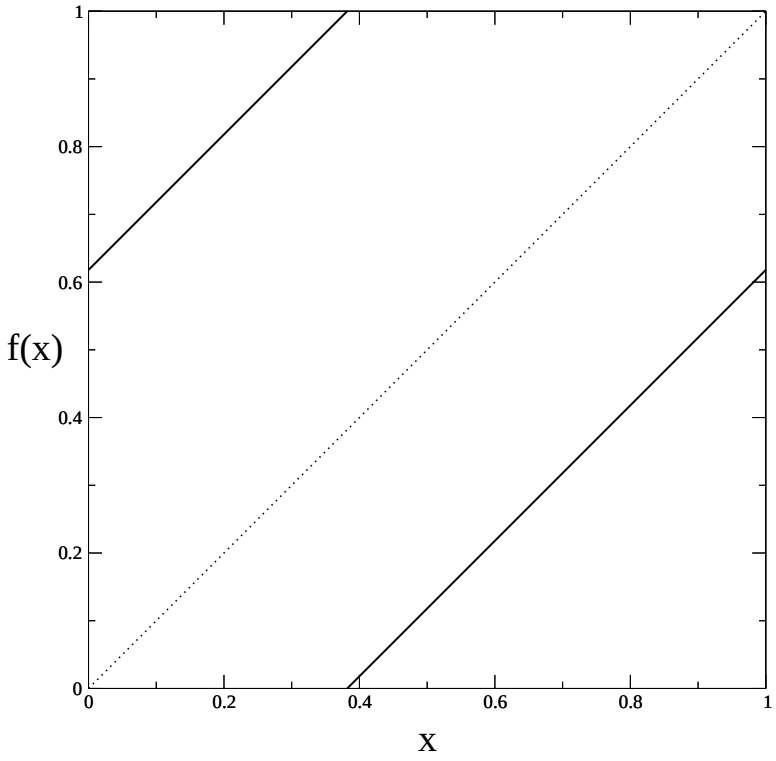


Figure 19.1: Unperturbed circle map ($k = 0$ in (19.1)) with golden mean rotation number.

and the rotation number is given by the parameter Ω.

$$W(k = 0, \Omega) = \Omega \ .$$

For given values of Ω and k the winding number can be either rational or irrational. For invertible maps and rational winding numbers $W = P/Q$ the asymptotic iterates of the map converge to a unique attractor, a stable periodic orbit of period Q

$$\hat{f}^Q(\hat{x}_i) = \hat{x}_i + P, \quad i = 0, 1, 2, \cdots, Q-1 \ .$$

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 Ω values for which the iterates of the circle map are attracted to the P/Q cycle. This interval is called the P/Q *mode-locked* (or *stability*) interval, and its width is given by

19.1 on p. 447

$$\Delta_{P/Q} = Q^{-2\mu_{P/Q}} = \Omega^{right}_{P/Q} - \Omega^{left}_{P/Q} \ . \tag{19.5}$$

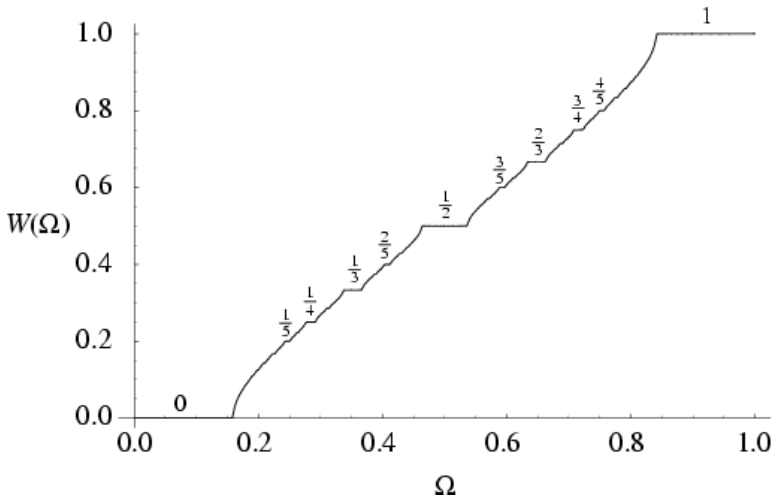


Figure 19.2: The critical circle map ($k = 1$ in (19.1)) devil's staircase [3]; the winding number W as function of the parameter Ω.

where $\Omega^{right}_{P/Q}$ ($\Omega^{left}_{P/Q}$) denote the biggest (smallest) value of Ω for which $W(k,\Omega) = P/Q$. Parametrizing mode lockings by the exponent μ rather than the width Δ will be convenient for description of the distribution of the mode-locking widths, as the exponents μ turn out to be of bounded variation. The stability of the P/Q cycle is

$$\Lambda_{P/Q} = \frac{\partial x_Q}{\partial x_0} = f'(x_0)f'(x_1)\cdots f'(x_{Q-1})$$

For a stable cycle $|\Lambda_{P/Q}|$ lies between 0 (the superstable value, the "center" of the stability interval) and 1 (the $\Omega^{right}_{P/Q}$, $\Omega^{left}_{P/Q}$ endpoints of (19.5)). For the shift map (19.4), the stability intervals are shrunk to points. As Ω is varied from 0 to 1, the iterates of a circle map either mode-lock, with the winding number given by a rational number $P/Q \in (0,1)$, or do not mode-lock, in which case the winding number is irrational. A plot of the winding number W as a function of the shift parameter Ω is a convenient visualization of the mode-locking structure of circle maps. It yields a monotonic "devil's staircase" of fig. 19.2 whose self-similar structure we are to unravel. Circle maps with zero slope at the inflection point x_c (see fig. 19.3)

$$f'(x_c) = 0\,, \qquad f''(x_c) = 0$$

($k = 1$, $x_c = 0$ in (19.1)) 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

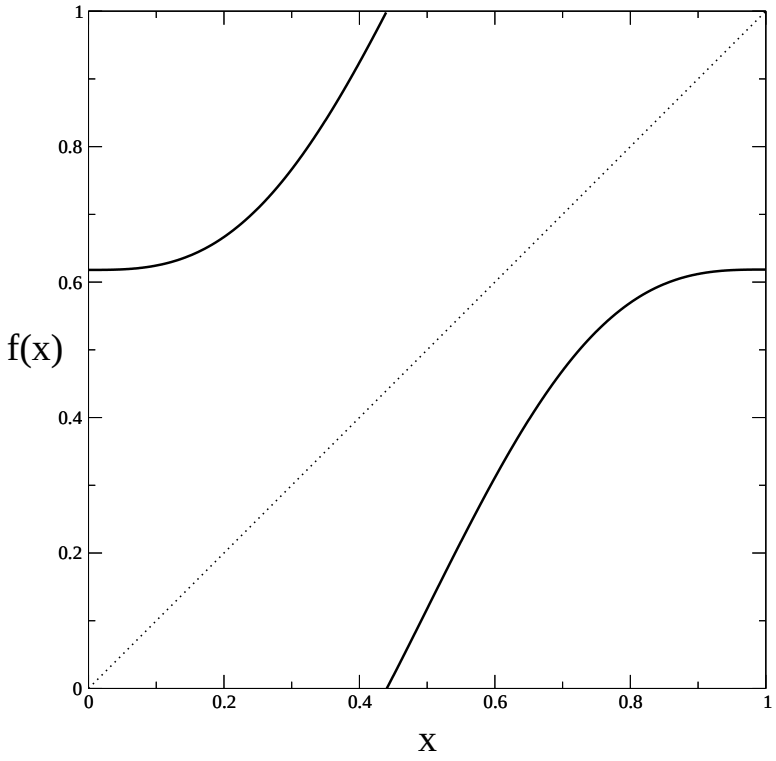


Figure 19.3: Critical circle map ($k = 1$ in (19.1)) with golden mean bare rotation number.

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.

19.1.1 Hierarchical partitions of the rationals

Intuitively, the longer the cycle, the finer the tuning of the parameter Ω required to attain it; given finite time and resolution, we expect to be able to resolve cycles up to some maximal 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 $0 < P_i \leq Q_i \leq Q$ and $(P_i|Q_i) = 1$. For example

$$\mathcal{F}_5 = \left\{ \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{1}{1} \right\}$$

A Farey series is characterized by the property that if P_{i-1}/Q_{i-1} and P_i/Q_i are consecutive terms of $\mathcal{F}_Q$, then

$$P_i Q_{i-1} - P_{i-1} Q_i = 1.$$

The number of terms in the Farey series F_Q is given by

$$\Phi(Q) = \sum_{n=1}^{Q} \phi(Q) = \frac{3Q^2}{\pi^2} + O(Q \ln Q). \tag{19.6}$$

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 nth 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 $\mathcal{S}_\infty = \{a_1 a_2 a_3 \cdots\}$, $a_k \in Z^+$, that is, the set of winding numbers with infinite continued fraction expansions. The continued fraction labeling is particularly appealing in the present context because of the close connection of the Gauss shift to the renormalization transformation R, discussed below. The Gauss map

$$T(x) = \begin{cases} \frac{1}{x} - \left[\frac{1}{x}\right] & x \neq 0 \\ 0\ , & x = 0 \end{cases} \tag{19.7}$$

($[\cdots]$ denotes the integer part) acts as a shift on the continued fraction representation of numbers on the unit interval

$$x = [a_1, a_2, a_3, \ldots] \ \to\ T(x) = [a_2, a_3, \ldots]\ . \tag{19.8}$$

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 fig. 19.2 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 Qth 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 nth *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

$$T_2 = \{[4], [2,2], [1,1,2], [1,3]\} = \left(\frac{1}{4}, \frac{1}{5}, \frac{3}{5}, \frac{3}{4}\right). \tag{19.9}$$

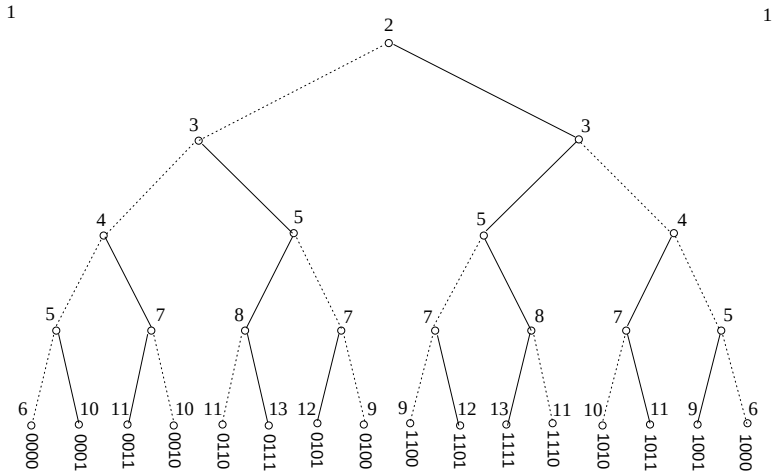


Figure 19.4: Farey tree: alternating binary ordered labelling of all Farey denominators on the nth Farey tree level.

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, labelling each by a unique binary label, fig. 19.4.

The smallest and the largest denominator in T_n are respectively given by

$$[n-2] \;=\; \frac{1}{n-2}\,, \qquad [1,1,\ldots,1,2] \;=\; \frac{F_{n+1}}{F_{n+2}} \propto \rho^n \;, \tag{19.10}$$

where the Fibonacci numbers F_n are defined by $F_{n+1} = F_n + F_{n-1}$; $\quad F_0 = 0$, $F_1 = 1$, and ρ is the golden mean ratio

$$\rho \;=\; \frac{1+\sqrt{5}}{2} \;=\; 1.61803\ldots \tag{19.11}$$

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.

19.2 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 α), in the parameter space (by shifting the parameter origin onto the border and rescaling by a factor δ), and by replacing the initial map f by the nth iterate f^n restricted to the magnified neighborhood

$$f_p(x) \to Rf_p(x) = \alpha f^n_{p/\delta}(x/\alpha)$$

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, that is the fix-point function of the renormalization operator R. The best known is the period doubling renormalization, with the recurrence times $n_i = 2^i$. The simplest circle map example is the golden mean renormalization, with recurrence times $n_i = F_i$ given by the Fibonacci numbers (19.10). Intuitively, in this context a metric self-similarity arises because iterates of critical maps are themselves critical, that is they also have cubic inflection points with vanishing derivatives.

The renormalization operator appropriate to circle maps acts as a generalization of the Gauss shift (19.38); 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 $[b, c, \ldots]$:

$$R_a \begin{pmatrix} g \\ f \end{pmatrix} = \begin{pmatrix} \alpha g^{a-1} \circ f \circ \alpha^{-1} \\ \alpha g^{a-1} \circ f \circ g \circ \alpha^{-1} \end{pmatrix}, \tag{19.12}$$

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 δ_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 δ_p.

For short repeating blocks, δ can be estimated numerically by comparing successive continued fraction approximants to W. Consider the P_r/Q_r rational approximation to a quadratic irrational winding number W_p whose continued

fraction expansion consists of r repeats of a block p. Let Ω_r be the parameter for which the map (19.1) has a superstable cycle of rotation number $P_r/Q_r = [p, p, \ldots, p]$. The δ_p can then be estimated by extrapolating from

$$\Omega_r - \Omega_{r+1} \propto \delta_p^{-r}. \tag{19.13}$$

What this means is that the "devil's staircase" of fig. 19.2 is self-similar under magnification by factor δ_p around any quadratic irrational W_p.

The fundamental result of the renormalization theory (and the reason why all this is so interesting) is that the ratios of successive P_r/Q_r mode-locked intervals converge to *universal* limits. The simplest example of (19.13) 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

$$\Delta_{1/Q} \propto Q^{-3} \tag{19.14}$$

For fixed Q the *widest* interval is bounded by $P/Q = F_{n-1}/F_n$, the nth 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

$$\Delta_{P/Q} \propto Q^{-2\mu_1} \tag{19.15}$$

where $P = F_{n-1}$, $Q = F_n$ and μ_1 is related to the universal Shenker number δ_1 (19.13) and the golden mean (19.11) by

$$\mu_1 = \frac{\ln|\delta_1|}{2\ln\rho} = 1.08218\ldots \tag{19.16}$$

The closeness of μ_1 to 1 indicates that the golden mean approximant mode-lockings barely feel the fact that the map is critical (in the k=0 limit this exponent is $\mu = 1$).

To summarize: for critical maps the spectrum of exponents arising from the circle maps renormalization theory is bounded from above by the harmonic scaling, and from below by the geometric golden-mean scaling:

$$3/2 > \mu_{m/n} \geq 1.08218\cdots. \tag{19.17}$$

19.3 Global theory: Thermodynamic averaging

Consider the following average over mode-locking intervals (19.5):

$$\Omega(\tau) = \sum_{Q=1}^{\infty} \sum_{(P|Q)=1} \Delta_{P/Q}^{-\tau}. \tag{19.18}$$

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 (19.18) is easy to pin down. For sufficiently negative τ, the sum is convergent; in particular, for $\tau = -1$, $\Omega(-1) = 1$, as for the critical circle maps the mode-lockings fill the entire Ω range [10]. However, as τ increases, the contributions of the narrow (large Q) mode-locked intervals $\Delta_{P/Q}$ get blown up to $1/\Delta_{P/Q}^{\tau}$, and at some critical value of τ the sum diverges. This occurs for $\tau < 0$, as $\Omega(0)$ equals the number of all rationals and is clearly divergent.

The sum (19.18) 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:

$$\bar{Z}_n(\tau) = \sum_{i \in \mathcal{S}_n} \Delta_i^{-\tau}. \tag{19.19}$$

The original sum (19.18) 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 (19.18):

$$\Omega(q,\tau) = \sum_{Q=1}^{\infty} \sum_{(P|Q)=1} e^{-q\nu_{P/Q}} Q^{2\tau\mu_{P/Q}} \ . \tag{19.20}$$

The sum (19.18) corresponds to $q = 0$. Exponents $\nu_{P/Q}$ will reflect the importance we assign to the P/Q mode-locking, that is 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

$$\Omega(q,\tau) = \sum_{Q=1}^{\infty} \Phi(Q)^{-q} \sum_{(P|Q)=1} Q^{2\tau\mu_{P/Q}} \ , \tag{19.21}$$

the continued fraction partitioning

$$\Omega(q,\tau) = \sum_{n=1}^{\infty} e^{-qn} \sum_{[a_1,\ldots,a_n]} Q^{2\tau\mu_{[a_1,\ldots,a_n]}} \ , \tag{19.22}$$

and the Farey tree partitioning

$$\Omega(q,\tau) = \sum_{k=n}^{\infty} 2^{-qn} \sum_{i=1}^{2^n} Q_i^{2\tau\mu_i} \ , \quad Q_i/P_i \in T_n \ . \tag{19.23}$$

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.

19.4 Hausdorff dimension of irrational windings

A finite cover of the set irrational windings at the "nth 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

$$\ell_i = \Omega^{min}_{P_r/Q_r} - \Omega^{max}_{P_l/Q_l} \ . \tag{19.24}$$

Here $\Omega^{min}_{P_r/Q_r}$ ($\Omega^{max}_{P_l/Q_l}$) are respectively the lower (upper) edges of the mode-locking intervals Δ_{P_r/Q_r} (Δ_{P_l/Q_l}) bounding ℓ_i and i is a symbolic dynamics label, for example the entries of the continued fraction representation $P/Q = [a_1, a_2, ..., a_n]$ of one of the boundary mode-lockings, $i = a_1 a_2 \cdots a_n$. ℓ_i provide a finite cover for the irrational winding set, so one may consider the sum

$$Z_n(\tau) \ = \ \sum_{i\in\mathcal{S}_n} \ell_i^{-\tau} \tag{19.25}$$

The value of $-\tau$ for which the $n \to \infty$ limit of the sum (19.25) is finite is the *Hausdorff dimension* D_H of the irrational winding set. Strictly speaking, this is the Hausdorff dimension only if the choice of covering intervals ℓ_i is optimal; otherwise it provides an upper bound to D_H. As by construction the ℓ_i intervals cover the set of irrational winding with no slack, we expect that this limit yields the Hausdorff dimension. This is supported by all numerical evidence, but a proof that would satisfy mathematicians is lacking.

The physically relevant statement is that for critical circle maps $D_H = 0.870\ldots$ is a (global) universal number. 19.2 on p. 447

19.4.1 The Hausdorff dimension in terms of cycles

Estimating the $n \to \infty$ limit of (19.25) from finite numbers of covering intervals ℓ_i is a rather unilluminating chore. Fortunately, there exist considerably more elegant ways of extracting D_H. We have noted that in the case of the "trivial" mode-locking problem (19.4), the covering intervals are generated by iterations of the Farey map (19.37) or the Gauss shift (19.38). The nth level sum (19.25) can be approximated by $\mathcal{L}_\tau^n$, where

$$\mathcal{L}_\tau(y,x) = \delta(x - f^{-1}(y))|f'(y)|^\tau$$

This amounts to approximating each cover width ℓ_i by $|df^n/dx|$ evaluated on the ith interval. We are thus led to the following determinant

$$\begin{aligned} \det(1 - z\mathcal{L}_\tau) &= \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{z^{rn_p}}{r} \frac{|\Lambda_p^r|^\tau}{1 - 1/\Lambda_p^r}\right) \\ &= \prod_p \prod_{k=0}^{\infty} \left(1 - z^{n_p}|\Lambda_p|^\tau/\Lambda_p^k\right) \ . \end{aligned} \tag{19.26}$$

The sum (19.25) 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 (19.26); putting all these pieces together, we obtain a pretty formula relating the Hausdorff dimension to the prime cycles of the map $f(x)$:

$$0 = \prod_p \left(1 - 1/|\Lambda_p|^{D_H}\right) \ . \tag{19.27}$$

For the Gauss shift (19.38) 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

$$x_a = \frac{-a + \sqrt{a^2+4}}{2}, \qquad \Lambda_a = -\left(\frac{a + \sqrt{a^2+4}}{2}\right)^2 \tag{19.28}$$

and the $x_{ab} = [a, b, a, b, a, b, \ldots]$ 2–cycles are given by

$$\begin{aligned} x_{ab} &= \frac{-ab + \sqrt{(ab)^2+4ab}}{2b} \\ \Lambda_{ab} &= (x_{ab}x_{ba})^{-2} = \left(\frac{ab+2+\sqrt{ab(ab+4)}}{2}\right)^2 \end{aligned} \tag{19.29}$$

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 (19.27) merely a very convoluted way to compute the number 1? Possibly so, but once the meaning of (19.27) has been grasped, the corresponding formula for the *critical* circle maps follows immediately:

$$0 = \prod_p \left(1 - 1/|\delta_p|^{D_H}\right) \ . \tag{19.30}$$

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 (19.30) are rather involved, as they require a heavy computational effort to extract Shenker's scaling δ_p for periodic continued fractions, and moreover dealing with an infinite alphabet requires control over tail summation if an accurate estimate is to be sought. In table 19.1 we give a small selection of computed Shenker's scalings.

19.5 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 15, by looking at the free energy.

p	δ_p
[1 1 1 1 ...]	-2.833612
[2 2 2 2 ...]	-6.7992410
[3 3 3 3 ...]	-13.760499
[4 4 4 4 ...]	-24.62160
[5 5 5 5 ...]	-40.38625
[6 6 6 6 ...]	-62.140
[1 2 1 2 ...]	17.66549
[1 3 1 3 ...]	31.62973
[1 4 1 4 ...]	50.80988
[1 5 1 5 ...]	76.01299
[2 3 2 3 ...]	91.29055

Table 19.1: Shenker's δ_p for a few periodic continued fractions, from ref. [1].

Consider the Farey tree partition sum (19.23): the narrowest mode-locked interval (19.15) at the nth level of the Farey tree partition sum (19.23) is the golden mean interval

$$\Delta_{F_{n-1}/F_n} \propto |\delta_1|^{-n}. \tag{19.31}$$

It shrinks exponentially, and for τ positive and large it dominates $q(\tau)$ and bounds $dq(\tau)/d\tau$:

$$q'_{max} = \frac{\ln|\delta_1|}{\ln 2} = 1.502642\ldots \tag{19.32}$$

However, for τ large and negative, $q(\tau)$ is dominated by the interval (19.14) which shrinks only harmonically, and $q(\tau)$ approaches 0 as

$$\frac{q(\tau)}{\tau} = \frac{3\ln n}{n\ln 2} \to 0. \tag{19.33}$$

So for finite n, $q_n(\tau)$ crosses the τ 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 τ for $-D_H \leq \tau$. This non-analyticity is rather severe - to get a clearer picture, we illustrate it by a few number-theoretic models (the critical circle maps case is qualitatively the same).

An approximation to the "trivial" Farey level thermodynamics is given by the "Farey model", in which the intervals $\ell_{P/Q}$ are replaced by Q^{-2}:

$$Z_n(\tau) = \sum_{i=1}^{2^n} Q_i^{2\tau}. \tag{19.34}$$

Here Q_i is the denominator of the ith Farey rational P_i/Q_i. For example (see fig. 19.4),

$$Z_2(1/2) = 4 + 5 + 5 + 4.$$

By the annihilation property (19.38) of the Gauss shift on rationals, the nth Farey level sum $Z_n(-1)$ can be written as the integral

$$Z_n(-1) = \int dx \delta(f^n(x)) = \sum 1/|f'_{a_1 \ldots a_k}(0)| \,,$$

and in general

$$Z_n(\tau) = \int dx \mathcal{L}^n_\tau(0,x) \,,$$

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 $x = 0$, that is by rationals, rather than by the quadratic irrationals as in (19.26). 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 19.2, they are consequence of the fact that the denominators on a given level are Farey sums of denominators on preceding levels.

19.3 on p. 447

A bound on D_H can be obtained by approximating (19.34) by

$$Z_n(\tau) = n^{2\tau} + 2^n \rho^{2n\tau}. \tag{19.35}$$

In this approximation we have replaced all $\ell_{P/Q}$, except the widest interval $\ell_{1/n}$, by the narrowest interval ℓ_{F_{n-1}/F_n} (see (19.15)). The crossover from the harmonic dominated to the golden mean dominated behavior occurs at the τ value for which the two terms in (19.35) contribute equally:

$$D_n = \hat{D} + O\left(\frac{\ln n}{n}\right), \quad \hat{D} = \frac{\ln 2}{2 \ln \rho} = .72\ldots \tag{19.36}$$

For negative τ the sum (19.35) is the lower bound on the sum (19.25) , so $\hat{D}$ is a lower bound on D_H.

$\tau/2$	$Z_n(\tau/2)/Z_{n-1}(\tau/2)$
0	2
1	3
2	$(5+\sqrt{17})/2$
3	7
4	$(5+\sqrt{17})/2$
5	$7+4\sqrt{6}$
6	26.20249...

Table 19.2: Partition function sum rules for the Farey model.

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 16 are important also in the analysis of renomarmalization at the onset of chaos.

Commentary

Remark **19.1** 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 Ω, and the angular motion can be modelled by a map periodic in the angular variable [6, 7]. 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 [8, 6] 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 **19.2** Critical mode–locking set The fact that mode-lockings completely fill the unit interval at the critical point has been proposed in refs. [3, 9]. The proof that the set of irrational windings is of zero Lebesgue measure in given in ref. [10].

Remark **19.3** 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 (19.25) of any methods used so far; for example the computation in ref. [11] for critical sine map (19.1), 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 **19.4** Farey tree presentation function. The Farey tree rationals can be generated by backward iterates of 1/2 by the Farey presentation function [12]:

$$\begin{array}{rcll} f_0(x) & = & x/(1-x) & 0 \leq x < 1/2 \\ f_1(x) & = & (1-x)/x & 1/2 < x \leq 1 \ . \end{array} \tag{19.37}$$

The Gauss shift (19.7) corresponds to replacing the binary Farey presentation function branch f_0 in (19.37) by an infinity of branches

$$\begin{array}{rcll} f_a(x) & = & f_1 \circ f_0^{(a-1)}(x) = \dfrac{1}{x} - a, & \dfrac{1}{a-1} < x \leq \dfrac{1}{a}, \\ f_{ab\cdots c}(x) & = & f_c \circ \cdot \circ f_b \circ f_a(x) \, . & \end{array} \tag{19.38}$$

A rational $x = [a_1, a_2, \ldots, a_k]$ is annihilated by the kth 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 (19.37) 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 phase 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. [13].

An alternative labelling of Farey denominators has been introduced by Knauf [14] 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 **19.5** Circle map renormalization The idea underlying golden mean renormalization goes back to Shenker [7]. A renormalization group

procedure was formulated in refs. [15, 16], where moreover the uniqueness of the relevant eigenvalue is claimed. This statement has been confirmed by a computer–assisted proof [17], and in the following we will always assume it. There are a number of experimental evidences for local universality, see refs. [18, 19].

On the other side of the scaling tale, the power law scaling for harmonic fractions (discussed in refs. [2, 3, 13]) is derived by methods akin to those used in describing intermittency [23]: $1/Q$ 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 **19.6** 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 [24, 25] . 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 [26]. A quantitive 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 [27, 24, 25], the Riemann hypothesis is equivalent to the statement that

$$\sum_{Q_i \le Q} |\delta_i| = o(Q^{\frac{1}{2}+\epsilon})$$

for all ϵ as $Q \to \infty$. The mode-lockings $\Delta_{P/Q}$ contain the necessary information for constructing the partition of the unit interval into the ℓ_i covers, and therefore implicitly contain the δ_i information. The implications of this for the circle-map scaling theory have not been worked out, and is not known whether some conjecture about the thermodynamics of irrational windings is equivalent to (or harder than) the Riemann hypothesis, but the danger lurks.

Remark **19.7** Farey tree partitioning. The Farey tree partitioning was introduced in refs. [28, 29, 13] and its thermodynamics is discussed in detail in refs. [11, 12]. 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 **19.8** Local and global universality. Numerical evidences for global universal behavior have been presented in ref. [3]. The question was reexamined in ref. [11], 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 (19.30), linking local to global behavior, was proposed in ref. [1].

The derivation of (19.30) relies only on the following aspects of the "hyperbolicity conjecture" of refs. [13, 20, 21, 22]:

1. *limits* for Shenker δ's *exist* and are universal. This should follow from the renormalization theory developed in refs. [15, 16, 17], though a general proof is still lacking.

2. δ_p grow *exponentially* with n_p, the length of the continued fraction block p.

3. δ_p for $p = a_1 a_2 \dots n$ with a large continued fraction entry n grows as a *power* of n. According to (19.14), $\lim_{n\to\infty} \delta_p \propto n^3$. 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 δ_p with n is not slower than a power of n.

Remark **19.9** Farey model. The Farey model (19.33) has been proposed in ref. [11]; 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 (19.26).

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.

References

[19.1] P. Cvitanović, G.H. Gunaratne and M. Vinson, *Nonlinearity* **3** (1990)

[19.2] K. Kaneko, *Prog. Theor. Phys.* **68**, 669 (1982); **69**, 403 (1983); **69**, 1427 (1983)

[19.3] 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)

[19.4] P. Cvitanović, B. Shraiman and B. Söderberg, *Physica Scripta* **32**, 263 (1985)

[19.5] J.A. Glazier and A. Libchaber, *IEEE Trans. Circ. Syst.*, **35**, 790 (1988)

[19.6] S.J. Shenker and L.P. Kadanoff, *J. Stat. Phys.* **27**, 631 (1982)

[19.7] S.J. Shenker, *Physica* **5D**, 405 (1982)

[19.8] J.M. Greene, *J. Math. Phys.* **20**, 1183 (1979)

[19.9] O.E. Lanford, *Physica* **14D**, 403 (1985)

[19.10] G. Swiatek, *Commun. Math. Phys.* **119**, 109 (1988)

[19.11] 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)

[19.12] M.J. Feigenbaum, *J.Stat.Phys.* **52**, 527 (1988)

[19.13] P. Cvitanović, B. Shraiman and B. Söderberg, *Physica Scripta* **32**, 263 (1985)

[19.14] A. Knauf, "On a ferromagnetic spin chain", *Commun. Math. Phys.* **153**, 77 (1993).

[19.15] M.J. Feigenbaum, L.P. Kadanoff, S.J. Shenker, *Physica* **5D**, 370 (1982)

[19.16] S. Ostlund, D.A. Rand, J. Sethna and E. Siggia, *Physica* **D 8**, 303 (1983)

[19.17] B.D. Mestel, Ph.D. Thesis (U. of Warwick 1985).

[19.18] J. Stavans, F. Heslot and A. Libchaber, *Phys. Rev. Lett.* **55**, 569 (1985)

[19.19] E.G. Gwinn and R.M. Westervelt, *Phys. Rev. Lett.* **59**, 157 (1987)

[19.20] 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)

[19.21] S.-H. Kim and S. Ostlund, *Physica* **D 39**, 365, (1989)

[19.22] M.J. Feigenbaum, *Nonlinearity* **1**, 577 (1988)

[19.23] Y. Pomeau and P. Manneville, *Commun. Math. Phys.* **74**, 189 (1980); P. Manneville, *J. Phys.* (Paris) **41**, 1235 (1980)

[19.24] H.M. Edwards, *Riemann's Zeta Function* (Academic, New York 1974)

[19.25] E.C. Titchmarsh, *The Theory of Riemann Zeta Function* (Oxford Univ. Press, Oxford 1951); chapter XIV.

[19.26] E.H. Neville, *Roy. Soc. Mathematical Tables* (Cambridge U. Press, Cambridge 1950)

[19.27] J. Franel and E. Landau, *Göttinger Nachr. 198* (1924)

[19.28] G. T. Williams and D. H. Browne, *Amer. Math. Monthly* **54**, 534 (1947)

[19.29] P. Cvitanović and J. Myrheim, *Phys. Lett.* **A94**, 329 (1983); *Commun. Math. Phys.* **121**, 225 (1989)

[19.30] P. Contucci and A. Knauf, *Forum Math.* **9**, 547 (1997)

Exercises

19.1 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.

19.2 Bounds on Hausdorff dimension. By making use of the bounds (19.17) show that the Hausdorff dimension for critical mode lockings may be bounded by

$$2/3 \leq D_H \leq .9240\ldots$$

19.3 Farey model sum rules. Verify the sum rules reported in table 19.2. 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. [30].

Chapter 20

Statistical mechanics

RM 8sep98

(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.

20.1 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 β, as it is a finite sum of positive terms.

The thermodynamic limit is also of central importance in the study of field theories. A field theory can be first defined on a lattice and then the lattice spacing is taken to zero as the correlation length is kept fixed. This continuum limit corresponds to the thermodynamic limit. In lattice spacing units the correlation length is going to infinity, and the interacting field theory can be thought of as a statistical mechanics model at a phase transition.

For general systems the convergence towards the thermodynamic limit is slow. If the thermodynamic limit exists for an interaction, the convergence of the free energy per unit volume f is as an inverse power in the linear dimension of the system.

$$f(\beta) \to \frac{1}{n} \tag{20.1}$$

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 nth 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,

$$f(\beta) \to e^{-\alpha n} \tag{20.2}$$

and may only be faster than that if all long-range correlations of the system are zero — that is, when there are no interactions. The coefficient α depends only on the inverse correlation length of the system.

One of the difficulties in using the transfer matrix techniques is that they seem at first limited to systems with finite range interactions. Phase transitions can happen only when the interaction is long range. One can try to approximate the long range interaction with a series of finite range interactions that have an ever increasing range. The problem with this approach is that in a formally defined transfer matrix, not all the eigenvalues of the matrix correspond to eigenvalues of the system (in the sense that the rate of decay of correlations is not the ratio of eigenvalues).

Knowledge of the correlations used in conjunction with finite size scaling to obtain accurate estimates of the parameters of systems with phase transitions. (Accurate critical exponents are obtained by series expansions or transfer matrices, and seldomly by renormalization group arguments or Monte Carlo.) In a phase transition the coefficient α of the exponential convergence goes to zero and the convergence to the thermodynamic limit is power-law.

The computation of the partition function is an example of a functional integral. For most interactions these integrals are ill-defined and require some form of normalization. In the spin models case the functional integral is very simple, as "space" has only two points and only "time" being infinite has to be dealt with. The same problem occurs in the computation of the trace of transfer matrices of systems with infinite range interactions. If one tries to compute the partition function Z_n

$$Z_n = \operatorname{tr} T^n$$

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.

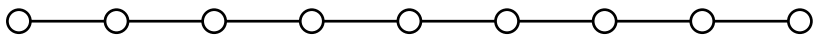
20.2 Ising models

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.

20.2.1 Ising model

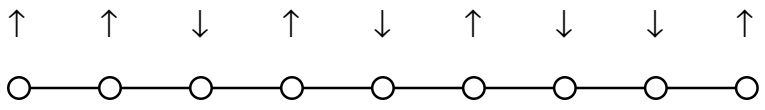
One of the simplest models in statistical mechanics is the Ising model. One imagines that one has a one-dimensional lattice with small magnets at each site that can point either up or down.

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



(20.3)

(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 ($\uparrow$) or down ($\downarrow$). This represents the two states of the small magnet on that site, and in general we will use the notation Σ_0 to represent the set of possible values of a spin at any site; all sites assume the same set of values. A configuration consists of assigning a value to the spin at each site; a typical configuration is



(20.4)

The set of all configurations for a lattice with n sites is called Ω_0^n and is formed by the Cartesian product $\Omega_0 \times \Omega_0 \cdots \times \Omega_0$, the product repeated n times. Each configuration $\sigma \in \Omega^n$ is a string of n spins

$$\sigma = \{\sigma_0, \sigma_1, \ldots \sigma_n\}\,, \tag{20.5}$$

In the example configuration (20.4) 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

$$H(\sigma) = \sum_i J\delta(\sigma_i, \sigma_{i+1})\,, \tag{20.6}$$

where

$$\delta(\sigma_1, \sigma_2) = \begin{cases} +1 & \text{if } \sigma_1 = \sigma_2 \\ -1 & \text{if } \sigma_1 \neq \sigma_2 \end{cases}\,. \tag{20.7}$$

One of the problems that was avoided when computing the energy was what to do at the boundaries of the one-dimensional chain. Notice that as written, (20.6) requires the interaction of spin n with spin $n+1$. In the absence of phase transitions the boundaries do not matter much to the thermodynamic limit and we will connect the first site to the last, implementing periodic boundary conditions.

Thermodynamic quantities are computed from the partition function $Z^{(n)}$ as the size n of the system becomes very large. For example, the free energy per site f at inverse temperature β is given by

$$-\beta f(\beta) = \lim_{n\to\infty} \frac{1}{n} \ln Z^{(n)}\,. \tag{20.8}$$

The partition function $Z^{(n)}$ is computed by a sum that runs over all the possible configurations on the one-dimensional chain. Each configuration contributes with its Gibbs factor $\exp(-\beta H(\sigma))$ and the partition function $Z^{(n)}$ is

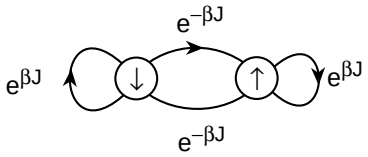
$$Z^{(n)}(\beta) = \sum_{\sigma\in\Omega_0^n} e^{-\beta H(\sigma)}\,. \tag{20.9}$$

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 frst thing to do is to expand out the sum for the energy of the configuration

$$Z^{(n)}(\beta) = \sum_{\sigma \in \Omega^n} e^{\beta J \delta(\sigma_1,\sigma_2)} e^{\beta J \delta(\sigma_2,\sigma_3)} \cdots e^{\beta J \delta(\sigma_n,\sigma_1)} \,. \tag{20.10}$$

Let us use the configuration in (20.4). The first site is $\sigma_1 = \uparrow$. As the second site is $\uparrow$, we know that the first term in (20.10) is a term $e^{\beta J}$. The third spin is $\downarrow$, so the second term in (20.10) is $e^{-\beta J}$. If the third spin had been $\uparrow$, then the term would have been $e^{\beta J}$ but it would not depend on the value of the first spin σ_1. This means that the configuration can be built site by site and that to compute the Gibbs factor for the configuration just requires knowing the last spin added. We can then think of the configuration as being a weighted random walk where each step of the walk contributes according to the last spin added. The random walk take place on the 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

$$T(\beta) = \begin{bmatrix} e^{\beta J} & e^{-\beta J} \\ e^{-\beta J} & e^{\beta J} \end{bmatrix} . \tag{20.11}$$

As in chapter **??** the sum of all closed walks is given by the trace of powers of the matrix. These powers can easily be re-expressed in terms of the two eigenvalues λ_1 and λ_2 of the transfer matrix:

$$Z^{(n)}(\beta) = \operatorname{tr} T^n(\beta) = \lambda_1(\beta)^n + \lambda_2(\beta)^n \,. \tag{20.12}$$

20.2.2 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?

20.2.3 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 Ω_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 nth 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.

20.3 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 three-dimensions, the analytic solution of the model shows that it is equivalent to a one-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.

20.3.1 Solution

The droplet model is a one-dimensional lattice gas where each site can have two states: empty or occupied. We will represent the empty state by 0 and the occupied state by 1. The configurations of the model in this notation are then strings of zeros and ones. Each configuration can be viewed as groups of contiguous ones separated by one or more zeros. The contiguous ones represent the droplets in the model. The droplets do not interact with each other, but the individual particles within each droplet do.

To determine the thermodynamics of the system we must assign an energy to every configuration. At very high temperatures we would expect a gaseous phase where there are many small droplets, and as we decrease the temperature the droplets would be expected to coalesce into larger ones until at some point there is a phase transition and the configuration is dominated by one large drop. To construct a solvable model and yet one with a phase transition we need long range interaction among all the particles of a droplet. One choice is to assign a fixed energy θ_n for the interactions of the particles of a cluster of size n. In a given droplet one has to consider all the possible clusters formed by contiguous particles. Consider for example the configuration 0111010. It has two droplets, one of size three and another of size one. The droplet of size one has only one cluster of size one and therefore contributes to the energy of the configuration with θ_1. The cluster of size three has one cluster of size three, two clusters of size two, and three clusters of size one; each cluster contributing a θ_n term to the energy. The total energy of the configuration is then

$$H(0111010) = 4\theta_1 + 2\theta_2 + 1\theta_3 \,. \tag{20.13}$$

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

$$a = \sum_{n>0} \theta_n < \infty \,. \tag{20.14}$$

The configuration with all zeros does not contribute to the energy.

Once we specify the function θ_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

$$G(z,\beta) = \sum_{n>0} z^n \frac{Z_n(\beta)}{n} \tag{20.15}$$

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

$$G(z,\beta) = \sum_{b\in B} \frac{z^{|b|}}{s(b)} e^{-\beta H(b)} \,. \tag{20.16}$$

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

(20.17)

Each droplet of size n contributes with energy

$$W_n = \sum_{1\le k\le n} (n-k+1)\theta_k \,. \tag{20.18}$$

So if we consider the sum

$$\sum_{n\ge 1} \frac{1}{n} \Big(z^2 e^{-\beta H(01)} + z^3 (e^{-\beta H(001)} + e^{-\beta H(011)}) + \\ + z^4 (e^{-\beta H(0001)} + e^{-\beta H(0011)} + e^{-\beta H(0111)}) + \cdots \Big)^n \tag{20.19}$$

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

$$-\ln\Big(1 - (z^2 e^{-\beta W_1} + z^3 (e^{-\beta W_1} + e^{-\beta W_2}) + \cdots\Big) \,, \tag{20.20}$$

where the $H(b)$ factors have been evaluated in terms of the droplet energies W_n. A proof of the equality of (20.19) and (20.20) 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

$$-\ln\Big(1 - (z^1 + z^2 + z^3 + \cdots)(z e^{-\beta W_1} + z^2 e^{-\beta W_2} + \cdots)\Big) \tag{20.21}$$

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

$$\sum_{n>0} \frac{z^n}{n} \,. \tag{20.22}$$

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

$$\sum_{n>0} \frac{z^n e^{-\beta H(\overbrace{11\ldots11}^{n})}}{n} \,. \tag{20.23}$$

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 (20.21), (20.22), and (20.23) we can evaluate the generating function G to be

$$G(z,\beta) = -\ln(1-z) - \ln(1-ze^{-\beta a}) - \ln(1-\frac{z}{1-z}\sum_{n\geq 1} z^n e^{-\beta W_n}) \,. \tag{20.24}$$

The cycle expansion $\zeta^{-1}(z,\beta)$ is given by the exponential of the generating function e^{-G} and we obtain

$$\zeta^{-1}(z,\beta) = (1-ze^{-\beta a})(1-z(1+\sum_{n\geq 1} z^n e^{-\beta W_n})) \tag{20.25}$$

To pursue this model further we need to have some assumptions about the interaction strengths θ_n. We will assume that the interaction strength decreases with the inverse square of the size of the cluster, that is, $\theta_n = -1/n^2$. With this we can estimate that the energy of a droplet of size n is asymptotically

$$W_n \sim -n + \ln n + \mathcal{O}(\frac{1}{n}) \,. \tag{20.26}$$

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 $\mathcal{O}$ 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

$$W_n = -n + \ln n \tag{20.27}$$

in a system of units where the dimensional constants are one. To evaluate the cycle expansion (20.25) we need to evaluate the constant a, the sum of all the θ_n. One can write a recursion for the θ_n

$$\theta_n = W_n - \sum_{1\le k<n} (n-k+1)\theta_k \tag{20.28}$$

and with an initial choice for θ_1 evaluate all the others. It can be verified that independent of the choice of θ_1 the constant a is equal to the number that multiplies the n term in (20.27). In the units used

$$a = -1\,. \tag{20.29}$$

For the choice of droplet energy (20.27) the sum in the cycle expansion can be expressed in terms of a special function: the Lerch transcendental ϕ_L. It is defined by

$$\phi_L(z,s,c) = \sum_{n\ge 0} \frac{z^n}{(n+c)^s}\,, \tag{20.30}$$

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 (20.25) using (20.27) as

$$\zeta^{-1}(z,\beta) = \left(1-ze^{\beta}\right)\left(1-z(1+\phi_L(ze^{\beta},\beta,1))\right) \tag{20.31}$$

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 (20.31). The root can come from any of the two factors. For large value of β (low temperatures) the smallest root is determined from the $(1-ze^{\beta})$ factor, which gave the contribution of a single large drop. For small β (large temperatures) the root is determined by the zero of the other factor, and it corresponds to the contribution from the gas phase of the droplet model. The transition occurs when the smallest root of each of the factors become numerically equal. This determines the critical temperature β_c through the equation

$$1 - e^{-\beta_c}(1+\zeta_R(\beta_c)) = 0 \tag{20.32}$$

which can be solved numerically. One finds that $\beta_c = 1.40495$. The phase transition occurs because the roots from two different factors get swapped in their roles as the smallest root. This in general leads to a first order phase transition. For large β the Lerch transcendental is being evaluated at the branch point, and therefore the cycle expansion cannot be an analytic function at low temperatures. For large temperatures the smallest root is within the radius of convergence of the series for the Lerch transcendental, and the cycle expansion has a domain of analyticity containing the smallest root.

As we approach the phase transition point as a function of β the smallest root and the branch point get closer together until at exactly the phase transition they collide. This is a sufficient condition for the existence of a first order phase transitions. In the literature of zeta functions [23] there have been speculations on how to characterize a phase transition within the formalism. The solution of the Fisher droplet model suggests that for first order phase transitions the factorized cycle expansion will have its smallest root within the radius of convergence of one of the series except at the phase transition when the root collides with a singularity. This does not seem to be the case for second order phase transitions.

The analyticity of the cycle expansion can be restored if we consider separate cycle expansions for each of the phases of the system. If we separate the two terms of ζ^{-1} in (20.31), each of them is an analytic function and contains the smallest root within the radius of convergence of the series for the relevant β values.

20.4 Scaling functions

> "Clouds are not spheres, mountains are not cones, coastlines are not circles and bark is not smooth, nor does lightning travel in straight line."
>
> B.B. Mandelbrot

There is a relation between general spin models and dynamical system. If one thinks of the boxes of the Markov partition of a hyperbolic system as the states of a spin system, then computing averages in the dynamical system is carrying out a sum over all possible states. One can even construct the natural measure of the dynamical system from a translational invariant "interaction function" call the scaling function.

There are many routes that lead to an explanation of what a scaling function is and how to compute it. The shortest is by breaking away from the historical development and considering first the presentation function of a fractal. The presentation function is a simple chaotic dynamical system (hyperbolic, unlike the circle map) that generates the fractal and is closely related to the definition

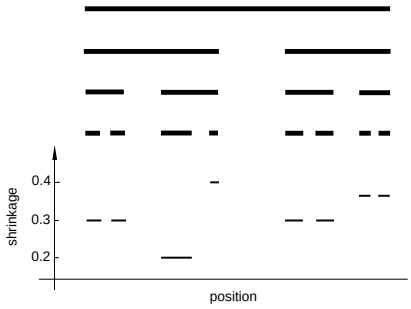


Figure 20.1: Construction of the steps of the scaling function from a Cantor set. From one level to the next in the construction of the Cantor set the covers are shrunk, each parent segment into two children segments. The shrinkage of the last level of the construction is plotted and by removing the gaps one has an approximation to the scaling function of the Cantor set.

of fractals of Hutchinson [24] and the iterated dynamical systems introduced by Barnsley and collaborators [20]. 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 fig. 20.1 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 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 fig. 20.2. Drawn is the map

$$F(x) = 5x(1-x) \tag{20.33}$$

restricted to the unit interval. We will see that this map is also a presentation function.

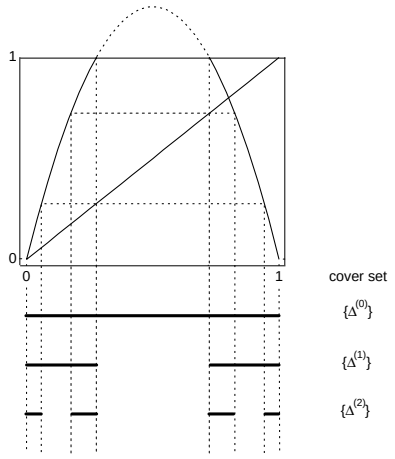


Figure 20.2: A Cantor set presentation function. The Cantor set is the set of all points that under iteration do not leave the interval $[0,1]$. This set can be found by backwards iterating the gap between the two branches of the map. The dotted lines can be used to find these backward images. At each step of the construction one is left with a set of segments that form a cover of the Cantor set.

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, (20.33), 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 $\mathcal{C}$ (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 $\mathcal{C}$, and has been called a presentation function by Feigenbaum [12].

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 fig. 20.2. 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 nth 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 fig. 20.2. As $n \to \infty$ the segments that remain for at least n iterates converge to the Cantor set $\mathcal{C}$.

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 fig. 20.2 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

$$\Delta^{(2)}(0,1) = F_0^{-1}\left(F_1^{-1}\left(\Delta^{(0)}\right)\right) . \tag{20.34}$$

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. In sect. **??** we will describe the limiting process in more detail, and give a precise definition on how to arrange the ratios.

20.5 Geometrization

The $\mathcal{L}$ operator is a generalization of the transfer matrix. It gets more by considering less of the matrix: instead of considering the whole matrix it is possible to consider just one of the rows of the matrix. The $\mathcal{L}$ operator also makes explicit the vector space in which it acts: that of the observable functions. Observables are functions that to each configuration of the system associate a number: the energy, the average magnetization, the correlation between two sites. It is in the average of observables that one is interested in. Like the transfer matrix, the $\mathcal{L}$ operator considers only semi-infinite systems, that is, only the part of the interaction between spins to the right is taken into account. This may sound un-symmetric, but it is a simple way to count each interaction only once, even in cases where the interaction includes three or more spin couplings. To define the $\mathcal{L}$ operator one needs the interaction energy between one spin and all the rest to its right, which is given by the function ϕ. The $\mathcal{L}$ operators defined as

$$\mathcal{L}g(\sigma) = \sum_{\sigma_0 \in \Omega_0} g(\sigma_0 \sigma) e^{-\beta \phi(\sigma_0 \sigma)} \,.$$

To each possible value in Ω_0 that the spin σ_0 can assume, an average of the observable g is computed weighed by the Boltzmann factor $e^{-\beta\phi}$. The formal relations that stem from this definition are its relation to the free energy when applied to the observable ι that returns one for any configuration:

$$-\beta f(\beta) = \lim_{n\to\infty} \frac{1}{n} \ln \|\mathcal{L}^n \iota\|$$

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_{σ_k} such that

$$F_{\sigma_k}(0) = 0$$

and

$$F_{\sigma_1} \circ F_{\sigma_2} \circ \cdots \circ F_{\sigma_n}(0) = \phi(+, \sigma_1, \sigma_2, \ldots, \sigma_n, -, -, \ldots)\,.$$

This is a formal relation that expresses how the interaction is to be converted into a dynamical systems. In most examples F_{σ_k} is a collection of maps from a subset of R^D to itself.

If the interaction is complicated, then the dimension of the set of maps may be infinite. If the resulting dynamical system is infinite have we gained anything from the transformation? The gain in this case is not in terms of added speed of convergence to the thermodynamic limit, but in the fact that the Ruelle operator is of trace-class and all eigenvalues are related to the spin system and not artifacts of the computation.

The construction of the higher dimensional system is done by borrowing the phase space reconstruction technique from dynamical systems. Phase 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 bescause 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 rrange 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 differnt 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 Ω. Then an observable g is a function from the set of configurations Ω to the reals. Each configuration is indexed by the integers from 0 up, and it is useful to think of the configuration as a string of spins. One can append a spin η_0 to its begining, $\eta \vee \sigma$, in which case η is at site 0, ω_0 at site 1, and so on.

The Ruelle operator $\mathcal{L}$ is defined as

$$\mathcal{L}g(\eta) = \sum_{\omega_0 \in \Omega_0} g(\omega_0 \vee \eta) e^{-\beta\phi(\omega_0 \vee \eta)} .$$

This is a positive and bounded operator over the space of bounded observables. There is a generalization of the Perron-Frobenius theorem by Ruelle that establishes that the largest eigenvalue of $\mathcal{L}$ is isolated from the rest of the spectrum and gives the thermodynamics of the spin system just as the largest eigenvalue of the transfer matrix does. Ruelle alos 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 nth power, $\operatorname{tr} \mathcal{L}^n = Z_n$ 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 σ the configuration can be reconstructed. We also introduce the interaction observables h_{σ_0}

To geometrize spin systems, the interactions are assumed to be translationally invariant. The spins σ_k will only assume a finite number of values. For simplicity, we will take the interaction ϕ among the spins to depend only on pairwise interactions,

$$\phi(\sigma) = \phi(\sigma_0, \sigma_1, \sigma_2, \ldots) = J_0\sigma_0 + \sum_{n>0} \delta_{\sigma_0,\sigma_n} J_1(n)\,, \tag{20.35}$$

and limit σ_k to be in $\{+,-\}$. For the one-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 three-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 nth 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 $\sigma = \{\sigma_0, \sigma_1, \ldots\}$:

$$\mathcal{T}g(\sigma) = g(+ \vee \sigma)e^{-\beta\phi(+\vee\sigma)} + g(- \vee \sigma)e^{-\beta\phi(-\vee\sigma)}\,. \tag{20.36}$$

By $+ \vee \sigma$ we mean the configuration obtained by prepending $+$ to the beginning of σ 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

$$-\beta f(\beta) = \lim_{n\to\infty} \frac{1}{n} \ln \|\mathcal{T}^n u\| \,. \tag{20.37}$$

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 σ, 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

$$\mathcal{T} G\left(x(\sigma)\right) = \sum_{\eta\in\{+,-\}} G\left(x_+\left(\eta \vee \sigma\right), x_-\left(\eta \vee \sigma\right)\right) e^{-\beta x_\eta(\sigma)} \,. \tag{20.38}$$

In this equation the only reference to the configuration σ is when computing the new observable values $x_+(\eta \vee \sigma)$ and $x_-(\eta \vee \sigma)$. The iteration of the function that gives these values in terms of $x_+(\sigma)$ and $x_-(\sigma)$ is the dynamical system that will reproduce the properties of the spin system. For the simple exponentially decaying potential this is given by two maps, F_+ and F_-. The map F_+ takes $\{x_+(\sigma), x_+(\sigma)\}$ into $\{x_+(+ \vee \sigma), x_-(+ \vee \sigma)\}$ which is $\{a(1 + x_+), ax_-\}$ and the map F_- takes $\{x_+, x_-\}$ into $\{ax_+, a(1 + x_-)\}$. In a more general case we have maps F_η that take $x(\sigma)$ to $x(\eta \vee \sigma)$.

We can now define a new operator $\mathcal{L}$

$$\mathcal{L} G\left(x\right) \stackrel{\text{def}}{=} \mathcal{T} G(x(\sigma)) = \sum_{\eta\in\{+,-\}} G\left(F_\eta(x)\right) e^{-\beta x_\eta} \,, \tag{20.39}$$

where all dependencies on σ have disappeared — if we know the value of the state observables x, the action of $\mathcal{L}$ on G can be computed.

A dynamical system is formed out of the maps F_η. They are chosen so that one of the state variables is the interaction energy. One can consider the two maps F_+ and F_- as the inverse branches of a hyperbolic map f, that is, $f^{-1}(x) = \{F_+(x), F_-(x)\}$. Studying the thermodynamics of the interaction ϕ is equivalent to studying the long term behavior of the orbits of the map f, achieving the transformation of the spin system into a dynamical system.

Unlike the original transfer operator, the $\mathcal{L}$ operator — acting in the space of observables that depend only on the state variables — is of trace-class (its trace is finite). The finite trace gives us a chance to relate the trace of $\mathcal{L}^n$ to the partition function of a system of size n. We can do better. As most properties of interest (thermodynamics, fall-off of correlations) are determined directly from its spectrum, we can study instead the zeros of the Fredholm determinant $\det(1-z\mathcal{L})$ by the technique of cycle expansions developed for dynamical systems [1]. A cycle expansion consists of finding a power series expansion for the determinant by writing $\det(1 - z\mathcal{L}) = \exp(\operatorname{tr}\ln(1 - z\mathcal{L}))$. The logarithm is expanded into a power series and one is left with terms of the form $\operatorname{tr}\mathcal{L}^n$ to evaluate. For evaluating the trace, the $\mathcal{L}$ operator is equivalent to

$$\mathcal{L}G(x) = \int_{\mathbf{R}^D} dy\, \delta(y - f(x))e^{-\beta y}G(y) \tag{20.40}$$

from which the trace can be computed:

$$\operatorname{tr}\mathcal{L}^n = \sum_{x=f^{(\circ n)}(x)} \frac{e^{-\beta H(x)}}{\left|\det\left(1 - \partial_x f^{(\circ n)}(x)\right)\right|} \tag{20.41}$$

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 $\operatorname{tr}\mathcal{L}^n$ 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 ϵ, such as $x = \phi(+, \epsilon_0, \epsilon_1, \ldots, -, -, \ldots)$. By choosing the sequence ϵ to have a large sequence of spins $-$, the number x can be made as small as needed, so in particular we can represent a small variation by $\phi(\eta)$. As $x_+(\epsilon) = \phi(+ \vee \epsilon)$, from the definition of a derivative we have:

$$\partial_x f(x) = \lim_{m\to\infty} \frac{\phi(\epsilon \vee \eta^{(m)}) - \phi(\epsilon)}{\phi(\eta^{(m)})}\,, \tag{20.42}$$

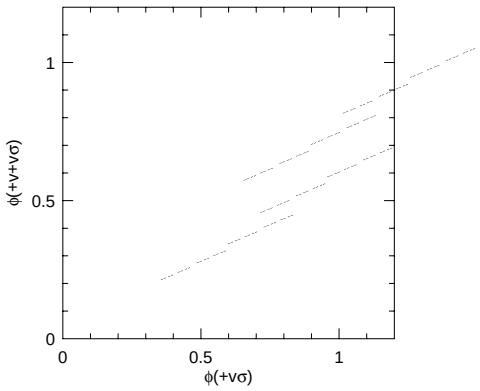


Figure 20.3: The spin adding map F_+ for the potential $J(n) = \sum n^2 a^{\alpha n}$. The action of the map takes the value of the interaction energy between $+$ and the semi-infinite configuration $\{\sigma_1, \sigma_2, \sigma_3, \ldots\}$ and returns the interaction energy between $+$ and the configuration $\{+, \sigma_1, \sigma_2, \sigma_3, \ldots\}$.

where $\eta^{(m)}$ is a sequence of spin strings that make $\phi(\eta^{(m)})$ smaller and smaller. By substituting the definition of ϕ in terms of its pairwise interaction $J(n) = n^s a^{n^\gamma}$ and taking the limit for the sequences $\eta^{(m)} = \{+, -, -, \ldots, \eta_{m+1}, \eta_{m+2}, \ldots\}$ 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

$$\phi(\sigma) = \sum_{n \geq 1} \delta_{\sigma_0, \sigma_n} \sum_k p_k(n) a_k^{n^\gamma} \tag{20.43}$$

where p_k are polynomials and $|a_k| < 1$, then the state observables to use are $x_{s,k}(\sigma) = \sum \delta_{+,\sigma_n} n^s a_k^n$. For each k one uses $x_{0,k}, x_{1,k}, \ldots$ 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 fig. 20.3. 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 [**?**, 16]. We will illustrate this with the polynomial-exponential interactions (20.43) with $\gamma = 1$, as the convergence is certainly faster than exponential

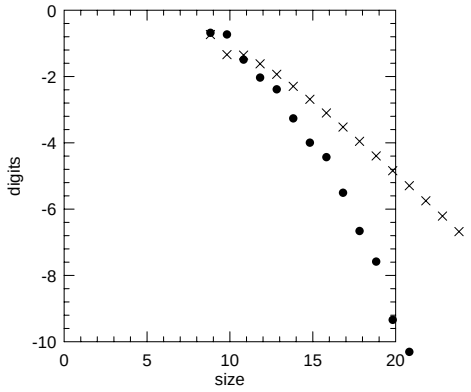


Figure 20.4: Number of digits for the Fredholm method (•) and the transfer function method (×). The size refers to the largest cycle considered in the Fredholm expansions, and the truncation length in the case of the transfer matrix.

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 fig. 20.4 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 $\mathrm{tr}\,\mathcal{L}^n$ 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 [**?**] 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.

Commentary

Remark **20.1** 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. [4]). Feigenbaum's *Journal of Statistical Physics* article [12] is the easiest place to learn about presentation functions.

Remark **20.2** 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 speedups 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 interac-

tion 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 δ are arbitrary discrete functions. This includes external constant fields (J_0), but it excludes site dependent fields such as a random external magnetic field.

Résumé

The geometrization of spin systems strengthens the connection between statistical mechanics and dynamical systems. It also further establishes the value of the Fredholm determinant of the $\mathcal{L}$ operator as a practical computational tool with applications to chaotic dynamics, spin systems, and semiclassical mechanics. The example above emphasizes the high accuracy that can be obtained: by computing the shortest 14 periodic orbits of period 5 or less it is possible to obtain three digit accuracy for the free energy. For the same accuracy with a transfer matrix one has to consider a 256×256 matrix. This make the method of cycle expansions practical for analytic calculations.

References

[20.1] Ya. Sinai. Gibbs measures in ergodic theory. *Russ. Math. Surveys*, 166:21–69, 1972.

[20.2] R. Bowen. Periodic points and measure for axiom-A diffeomorphisms. *Transactions Amer. Math. Soc.*, 154:377–397, 1971.

[20.3] D. Ruelle. Statistical mechanics on a compound set with Z^ν action satisfying expansiveness and specification. *Transactions Amer. Math. Soc.*, 185:237–251, 1973.

[20.4] E. B. Vul, Ya. G. Sinai, and K. M. Khanin. Feigenbaum universality and the thermodynamic formalism. *Uspekhi Mat. Nauk.*, 39:3–37, 1984.

[20.5] 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.

[20.6] 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.

[20.7] F. Takens, Detecting strange attractors in turbulence. In *Lecture Notes in Mathematics 898*, pages 366–381. Springer, Berlin, 1981.

[20.8] R. Mainieri. Thermodynamic zeta functions for Ising models with long range interactions. *Physical Review A*, 45:3580, 1992.

[20.9] R. Mainieri. Zeta function for the Lyapunov exponent of a product of random matrices. *Physical Review Letters*, 68:1965–1968, March 1992.

[20.10] D. Wintgen. Connection between long-range correlations in quantum spectra and classical periodic orbits. *Physical Review Letters*, 58(16):1589–1592, 1987.

[20.11] 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.

[20.12] H. A. Kramers and G. H. Wannier. Statistics of the two-dimensional ferromagnet. Part I. *Physical Review*, 60:252–262, 1941.

[20.13] D. Ruelle. Statistical mechanics of a one-dimensional lattice gas. *Communications of Mathematical Physics*, 9:267–278, 1968.

[20.14] David Ruelle. *Thermodynamic Formalism*. Addison-Wesley, Reading, 1978.

[20.15] P. Walters. *An introduction to ergodic theory*, volume 79 of *Graduate Text in Mathematics*. Springer-Verlag, New York, 1982.

[20.16] H.H. Rugh. *Time evolution and correlations in chaotic dynamical systems*. PhD thesis, Niels Bohr Institute, 1992.

[20.17] M.C. Gutzwiller. The quantization of a classically ergodic system. *Physica D*, 5:183–207, 1982.

[20.18] M. Feigenbaum. The universal metric properties of non-linear transformations. *Journal of Statistical Physics*, 19:669, 1979.

[20.19] G.A. Baker. One-dimensional order-disorder model which approaches a second order phase transition. *Phys. Rev.*, 122:1477–1484, 1961.

[20.20] M. E. Fisher. The theory of condensation and the critical point. *Physics*, 3:255–283, 1967.

[20.21] G. Gallavotti. Funzioni zeta ed insiemi basilari. *Accad. Lincei. Rend. Sc. fis. mat. e nat.*, 61:309–317, 1976.

[20.22] R. Artuso. Logarithmic strange sets. *J. Phys. A.*, 21:L923–L927, 1988.

[20.23] Dieter H. Mayer. *The Ruelle-Araki transfer operator in classical statistical mechanics*. Springer-Verlag, Berlin, 1980.

[20.24] Hutchinson

Exercises

20.1 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 ther is no scalar product that will induce this norm.

20.2 Automaton for a droplet Find the Markov graph and the weights on the edges so that the energies of configurations for the dropolet 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 (20.13) and (20.18).

Hint: the Markov graph is infinite.

20.3 Spectral determinant for a^n interactions Compute the spectral determinant for one-dimensional Ising model with the interaction

$$\phi(\sigma) = \sum_{k>0} a^k \delta(\sigma_0, \sigma_k)\,.$$

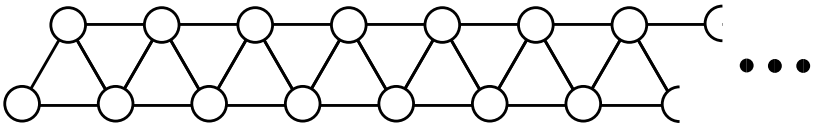
Take a as a number smaller than $1/2$.

(a) What is the dynamical system this generates? That is, find F_+ and F_- as used in (20.39).

(b) Show that

$$\frac{d}{dx}F_{\{+ \text{ or} -\}} = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$$

20.4 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.

20.5 Infinite symbolic dynamics Let σ be a function that returns zeo 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 ϵ_i are either 0 or 1. We will now define an operator $\mathcal{T}$ that acts on observables on the space of binary strings. A function a is an observable if it has bounded variation, that is, if

$$\|a\| = \sup_{\{\epsilon_i\}} |a(\epsilon_1, \epsilon_2, \ldots)| < \infty \,.$$

For these functions

$$\mathcal{T}a(\epsilon_1, \epsilon_2, \ldots) = a(0, \epsilon_1, \epsilon_2, \ldots)\sigma(0, \epsilon_1, \epsilon_2, \ldots) + a(1, \epsilon_1, \epsilon_2, \ldots)\sigma(1, \epsilon_1, \epsilon_2, \ldots) \,.$$

The function σ is assumed such that any of $\mathcal{T}$'s "matrix representations" in (a) have the Markov property (the matrix, if read as an adjacency graph, corresponds to a graph where one can go from any node to any other node).

(a) (easy) Consider a finite version T_n of the operator $\mathcal{T}$:

$$\begin{aligned} T_n a(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) = \\ a(0, \epsilon_1, \epsilon_2, \ldots, \epsilon_{n-1})\sigma(0, \epsilon_1, \epsilon_2, \ldots, \epsilon_{n-1}) + \\ a(1, \epsilon_1, \epsilon_2, \ldots, \epsilon_{n-1})\sigma(1, \epsilon_1, \epsilon_2, \ldots, \epsilon_{n-1}) \,. \end{aligned}$$

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.

Chapter 21

Semiclassical evolution

> Anyone who uses words "quantum" and "chaos" in the same sentence should be hung by his thumbs on a tree in the park behind the Niels Bohr Institute
> Joseph Ford

(G. Vattay, G. Tanner and P. Cvitanović)

So far we have obtained information about the nonlinear dynamics by calculating spectra of linear operators such as the Perron-Frobenius operator of sect. 21.2.3 or the associated partial differential equations such as the Liouville equation (5.35) of sect. 5.4.1. The spectra of these operators could then again be described in terms of the periodic orbits of the deterministic dynamics by means of trace formulas and cycle expansions. We also noted that the type of the dynamics has a strong influence on the convergence of cycle expansions and thus also on the spectra; this made it necessary to develop very 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 13 and 16.

We now address the converse question: 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 a 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, however not be exact, in contrast to the classical periodic orbit formulas obtained so far. Waves exhibit

new phenomena, such as interference, diffraction, and the higher $\hbar$ corrections which still need to be incorporated in the periodic orbit theory.

We will restrict the discussion in what follows to a wave equation of great import for physics, the non-relativistic Schrödinger equation. Our 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 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 answer was provided by 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, however, 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. We find instead that a coherent summation over the infinity of periodic orbit contributions gives the desired spectrum.

Before we get to this point we 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 short 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 22. In subsequent chapters we butress our case by applying and extyending the theory: a cycle expansion calculation of scattering resonances in a 3-disk billiard in chapter **??**, the spectrum of helium in chapter 23, and the incorporation of diffraction effects in chapter 24.

21.1 Quantum mechanics: A brief review

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

$$i\hbar\frac{\partial}{\partial t}\psi(q,t) = \hat{H}(q,\frac{\hbar}{i}\frac{\partial}{\partial q})\psi(q,t), \tag{21.1}$$

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

$$H(q,p) = T(p) + V(q)\,, \qquad T(p) = p^2/2m\,, \tag{21.2}$$

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

$$\psi_n(q,t) = e^{-iE_nt/\hbar}\phi_n(q), \tag{21.3}$$

where E_n are the eigenenergies of the system obtained by solving the eigenvalue equation $\hat{H}\phi_n(q) = E_n\phi_n(q)$. For bound systems the spectrum is discrete and the eigenfunctions form an orthonormal

$$\int d^dq\,\phi_n(q)\phi_m^*(q) = \delta_{nm} \tag{21.4}$$

and complete

$$\sum_n \phi_n(q)\phi_n^*(q') = \delta(q-q')\,. \tag{21.5}$$

set of functions in a Hilbert space. 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 ??

A given wave function can be expanded in the energy eigenbasis

$$\psi(q,t) = \sum_n c_n e^{-iE_nt/\hbar}\phi_n(q)\,, \tag{21.6}$$

where the expansion coefficient c_n is given by the projection of the initial wave function $\psi(q,0)$ onto the nth eigenstate

$$c_n = \int d^d q\, \phi_n^*(q)\psi(q,0). \tag{21.7}$$

By substituting (21.7) into (21.6), we can cast the evolution of a wave function into a multiplicative form

$$\psi(q,t) = \int d^d q' K(q,q',t)\psi(q',0)\,,$$

with the kernel

$$K(q,q',t) = \sum_n \phi_n(q)\, e^{-iE_n t/\hbar}\phi_n^*(q') \tag{21.8}$$

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

$$K(q,q',t_1+t_2) = \int dq''\, K(q,q'',t_2)K(q'',q',t_1) \tag{21.9}$$

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

$$i\hbar\frac{\partial}{\partial t}K(q,q',t) = \hat{H}(q, \frac{i}{\hbar}\frac{\partial}{\partial q})K(q,q',t)\,, \tag{21.10}$$

and is thus a wave function defined for $t \geq 0$; from the completeness relation (21.5) we obtain the boundary condition at $t = 0$:

$$\lim_{t\to 0_+} K(q,q',t) = \delta(q-q')\,. \tag{21.11}$$

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 soon as the eigenenergies E_n and eigenfunctions ϕ_n have been determined. With time dependence rendered "trivial", it makes sense to focus on obtained from the propagator via a Laplace transformation

$$G(q, q', E + i\epsilon) = \frac{1}{i\hbar} \int_0^\infty dt\, e^{\frac{i}{\hbar}Et - \frac{\epsilon}{\hbar}t} K(q, q', t) = \sum_n \frac{\phi_n(q)\phi_n^*(q')}{E - E_n + i\epsilon}\,. \quad (21.12)$$

Here ϵ is a small positive number, ensuring the existence of the integral. The eigenenergies show up as poles in the Green's function with residues corresponding to the wave function amplitudes. If one is only interested in the spectrum, one may restrict the considerations to the (formal) trace of the Green's function,

$$\mathrm{tr}\, G(q, q', E) = \int d^d q\, G(q, q, E) = \sum_n \frac{1}{E - E_n}\,, \quad (21.13)$$

where E is complex, with a positive imaginary part, and we have used the eigenfunction orthonormality (21.4). This trace is formal, since in general the sum in (21.13) is divergent. We shall return to this point in sects. 22.1.1 and 22.1.2.

A 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, fig. 21.1(a),

$$d(E) = \sum_n \delta(E - E_n). \quad (21.14)$$

Using the identity

21.1 on p. 507

$$\delta(E - E_n) = -\lim_{\epsilon \to +0} \frac{1}{\pi} \mathrm{Im}\, \frac{1}{E - E_n + i\epsilon} \quad (21.15)$$

we can express the density of states in terms of the trace of the Green's function, that is

$$d(E) = \sum_n \delta(E - E_n) = -\lim_{\epsilon \to 0} \frac{1}{\pi} \mathrm{Im}\, \mathrm{tr}\, G(q, q', E + i\epsilon). \quad (21.16)$$

sect. 22.1.1

This relation is the reason why we chose to describe the quantum spectrum in terms of the density of states. As we shall see, a semiclassical formula for right

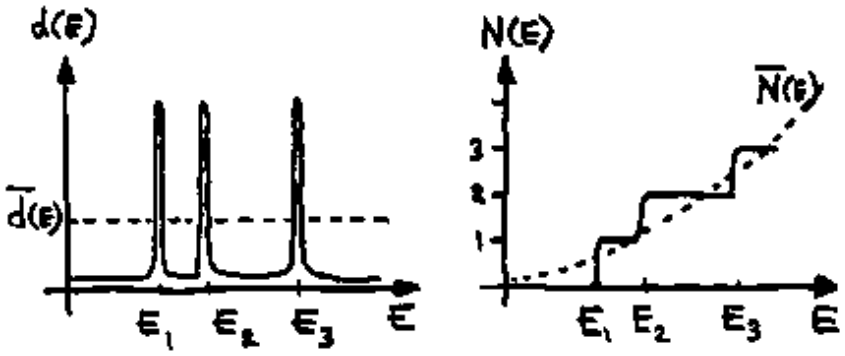


Figure 21.1: Schematic picture of **a)** the density of states $d(E)$, and **b)** the spectral staircase function $N(E)$. The dashed lines denote the mean density of states $\bar{d}(E)$ and the average number of states $\bar{N}(E)$ discussed in more detail in sect. 15.1.

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 another useful quantity, the *spectral staircase* function

$$N(E) = \sum_n \Theta(E - E_n) \tag{21.17}$$

which counts the number of eigenenergies below E, fig. 21.1(b). Here Θ is the Heaviside function

$$\Theta(x) = 1 \quad \text{if } x > 0; \quad \Theta(x) = 0 \quad \text{if } x < 0\,. \tag{21.18}$$

This completes our lightning review of quantum mechanics.

21.2 Semiclassical evolution

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 real novelty of quantum mechanics is the interference of the point-like particle with other versions of itself travelling along different classical trajectories, a feat impossible in classical mechanics. The short wavelength –

☞ remark 21.1

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.

The mathematical formulation of the semiclassical approximation starts out with a rewrite of the wave function

$$\psi(q,t) = A(q,t)e^{iR(q,t)/\hbar}\,, \tag{21.19}$$

in terms of a pair of real functions $R(q,t)$ and $A(q,t)$, its phase and magnitude. The time evolution of the phase and the magnitude of ψ follows from the Schrödinger equation (21.1)

21.2 on p. 507

$$\left(i\hbar\frac{\partial}{\partial t} + \frac{\hbar^2}{2m}\frac{\partial^2}{\partial q^2} - V(q)\right)\psi(q,t) = 0\,. \tag{21.20}$$

Take for concreteness a Hamiltonian $\hat{H}$ of form (21.2), 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

$$\frac{\partial R}{\partial t} + \frac{1}{2m}\left(\frac{\partial R}{\partial q}\right)^2 + V(q) - \frac{\hbar^2}{2m}\frac{1}{A}\frac{\partial^2}{\partial q^2}A = 0\,, \tag{21.21}$$

and the imaginary part giving the time evolution of the amplitude

21.3 on p. 507

$$\frac{\partial A}{\partial t} + \frac{1}{m}\sum_{i=1}^{d}\frac{\partial A}{\partial q_i}\frac{\partial R}{\partial q_i} + \frac{1}{2m}A\frac{\partial^2 R}{\partial q^2} = 0\,. \tag{21.22}$$

21.4 on p. 507

21.5 on p. 507

In this way a linear PDE for a complex wave function is converted into a set of coupled non-linear PDE's of real-valued functions R and A. The coupling term in (21.21) is, however, of order $\hbar^2$ and thus small in the semiclassical limit $\hbar \to 0$. Now we make the *Wentzel-Kramers-Brillouin* (WKB) *ansatz*: 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

$$\frac{\partial R}{\partial t} + H\left(q, \frac{\partial R}{\partial q}\right) = 0\,. \tag{21.23}$$

We recognize this as the classical *Hamilton-Jacobi equation*. We will refer to this step as the *semiclassical approximation* to wave mechanics, and from now on work only within this approximation.

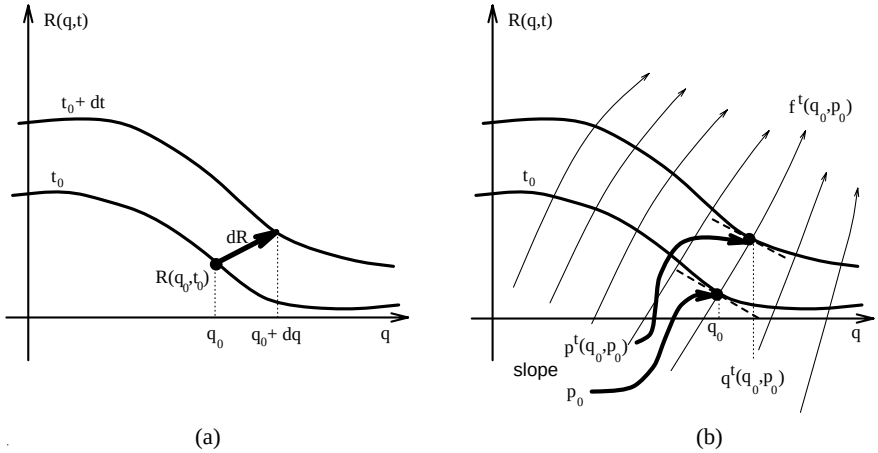


Figure 21.2: (a) A phase $R(q,t)$ plotted as a function of the position q for two infinitesimally close times. (b) The phase $R(q,t)$ transported by a swarm of "particles"; The Hamilton's equations (21.28) construct $R(q,t)$ by transporting $q_0 \to q(t)$ and the slope of $R(q_0,t_0)$, that is $p_0 \to p(t)$.

21.2.1 Hamilton's equations

> 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*

If you feel that you already understand the Hamilton-Jacobi theory, you can safely skip this section.

The wave equation (21.1) describes how the wave function ψ evolves with time, and if you think of ψ as an (infinite dimensional) vector, position q plays a 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 fig. 21.2(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, fig. 21.2(b). For notational convenience, define

$$p_i = p_i(q,t) := \frac{\partial R}{\partial q_i}\,, \quad i = 1,2,\ldots,d\,. \tag{21.24}$$

We saw earlier that (21.21) reduces in the semiclassical approximation to the Hamilton-Jacobi equation (21.23). To make life simple, we shall assume through-

out this chapter that the Hamilton's function $H(q,p)$ does not depend explicitly on time t, that is 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, fig. 21.2(a)

$$dR = \frac{\partial R}{\partial t}dt + \frac{\partial R}{\partial q}dq\,. \tag{21.25}$$

Dividing through by dt and substituting (21.23) we obtain the total derivative of $R(q,t)$ with respect to time *along as yet arbitrary direction* $\dot{q}$, that is,

$$\frac{dR}{dt}(q,\dot{q},t) = -H(q,p) + \dot{q}\cdot p\,. \tag{21.26}$$

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 (21.26) 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 (21.23) 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,

$$\dot{p}(q,\dot{q},t) + \frac{\partial H}{\partial q} = \left(\dot{q} - \frac{\partial H}{\partial p}\right)\frac{\partial p}{\partial q}\,. \tag{21.27}$$

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

$$\dot{q} = \frac{\partial H(q,p)}{\partial p}, \qquad \dot{p} = -\frac{\partial H(q,p)}{\partial q} \tag{21.28}$$

with initial conditions

$$q(t_0) = q', \qquad p(t_0) = p' = \frac{\partial R}{\partial q}(q', t_0). \tag{21.29}$$

☞ sect. 2.2.2

We recognize (21.28) as the Hamilton's equations of motion of classical mechanics.

$\dot{q}$ is no longer an independent function, and the phase $R(q,t)$ can now be computed by integrating equation (21.26) along the trajectory $(q(t), p(t))$

$$\begin{aligned} R(q,t) &= R(q',t_0) + R(q,t;q',t_0) \\ R(q,t;q',t_0) &= \int_{t_0}^{t} d\tau \left[\dot{q}(\tau)\cdot p(\tau) - H(q(\tau),p(\tau))\right] , \end{aligned} \tag{21.30}$$

with the initial conditions (21.29). In this way the Hamilton-Jacobi *partial* differential equation (21.21) is solved by integrating a set of *ordinary* differential equations, the 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 fig. 21.2(b). Erudite reader has already noticed that the integrand of (21.30) is known as the Lagrangian, and that a variational principle lurks somewhere, but we shall not make much fuss about that here.

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:

$$p' = \frac{\partial}{\partial q'} R(q',0) = -\frac{\partial}{\partial q'} R(q,q',t). \tag{21.31}$$

21.6 on p. 508

Function $R(q,q',t)$ is known as the *Hamilton's principal function*.

21.7 on p. 508

To summarize: Hamilton's achievement was to trade in the Hamilton-Jacobi *partial* differential equation (21.23) for the evolution of a wave front for a finite number of *ordinary* differential equations of motion which increment the initial phase $R(q,0)$ by the integral (21.30) along the phase space trajectory $(q(\tau), p(\tau))$.

21.2.2 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 (21.30) is simply Et. The first term, or the *action*

$$S(q,q',E) = \int_0^t d\tau\, \dot{q}(\tau)\cdot p(\tau) = \int_{q'}^{q} dq\cdot p \tag{21.32}$$

is integrated along a trajectory from q' to q with a given energy E. By (21.30) the action is the Legendre transform of Hamilton's principal function

$$S(q,q',E) = R(q,q',t) + Et\,. \tag{21.33}$$

The time of flight t along the trajectory connecting $q' \to q$ with fixed energy E is given by

$$\frac{\partial}{\partial E}S(q,q',E) = t\,. \tag{21.34}$$

The way to think about the formula (21.33) 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

$$\frac{\partial}{\partial t}R(q,q',t) = -E\,, \tag{21.35}$$

and the second variations of R and S are related in the standard way of Legendre transforms:

$$\frac{\partial^2}{\partial t^2}R(q,q',t)\frac{\partial^2}{\partial E^2}S(q,q',E) = -1\,. \tag{21.36}$$

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 ☞ sect. 21.2.4

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 contractable loop γ in it has zero action,

$$0 = \oint_\gamma dq \cdot p,$$

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 time t later. This fact is called the Poincaré-Cartan theorem.

Returning back to the main line of our argument: we show next that the velocity field given by the Hamilton's equations together with the continuity equation determines the amplitude of the wave function.

21.2.3 Density evolution

To obtain the full solution of the Schrödinger equation (21.1), we also have to integrate (21.22). Already Schrödinger noted that if one defines

$$\rho = \rho(q,t) := A^2 = \psi^*\psi$$

evaluated along the trajectory $(q(t), p(t))$, the amplitude equation (21.22) is equivalent to the continuity equation (5.34) after multiplying (21.22) by $2A$, that is

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial q_i}(\rho v_i) = 0\,. \tag{21.37}$$

Here, $v_i = \dot{q}_i = p_i/m$ denotes a velocity field, which is in turn determined by the phase $R(q,t)$ or equivalently by 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 (21.23), 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

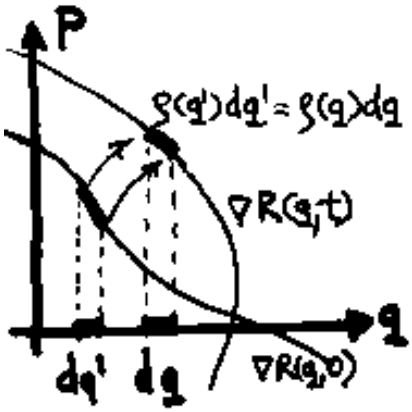


Figure 21.3: Density evolution of an initial surface $(q', p' = \partial_q R(q', 0)$ into $(q(t), p(t))$ surface time t later, sketched in 1 dimension. While the number of trajectories and the phase space Liouville volume are conserved, the density of trajectories projected on the q coordinate varies; trajectories which started in dq' at time zero end up in the interval dq.

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 $\rho(q, t)d^d q$. They started initially in a small volume $d^d q'$ around the point q' of the configuration space. For the moment, we assume that there is only one solution, the case of several paths will be considered below. The number of particles at time t in the volume is the same as the number of particles in the initial volume at $t = 0$,

$$\rho(q(t), t)d^d q = \rho(q', 0)d^d q' \,,$$

see fig. 21.3. The ratio of the initial and the final volumes can be expressed as

$$\rho(q(t), t) = \left|\det \mathbf{j}(q, q', t)\right| \rho(q', 0), \tag{21.38}$$

where $\mathbf{j}$ stands for the configuration space Jacobian ☞ sect. 5.2

$$\mathbf{j}(q, q', t) = \frac{\partial q'_l}{\partial q_k} \,. \tag{21.39}$$

As we know how to compute trajectories $(q(t), p(t))$, we know how to compute this jacobian and, by (21.38), the density $\rho(q(t), t)$ at time t.

21.2.4 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^dq configuration space volume element keeps its orientation. The evolved wave function is in the semiclassical approximation then given by

$$\begin{aligned}\psi_{sc}(q,t) &= A(q,t)e^{iR(q,t)/\hbar} = \sqrt{\det \mathbf{j}(q,q',t))}\, A(q',0)e^{i(R(q',0)+R(q,q',t))/\hbar} \\ &= \sqrt{\det \mathbf{j}(q,q',t)}\, e^{iR(q,q',t)/\hbar}\psi(q',0)\,.\end{aligned}$$

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 fig. 21.4.

Whenever the Lagrangian manifold develops a fold, the density of the phase space trajectories in the fold projected on the configuration coordinates diverges. Presence of a fold is signaled by the divergence of an eigenvalue of the Jacobian $\mathbf{j}$ from (21.39). 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 on the bottom of a swimming pool.

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.

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

$$\det \mathbf{j}^j(q,q',t) = e^{-i\pi m_j(q,q',t)}|\det \mathbf{j}^j(q,q',t)|,$$

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 fig. 21.4. We shall refer to the integer $m_j(q,q',t)$ as the *topological or Morse index* of the trajectory. So in general the semiclassical approximation to the wave function is thus a sum over possible trajectories that start in q' and end in q in time t

$$\psi_{sc}(q,t) = \sum_j |\det \mathbf{j}^j(q,q',t)|^{1/2} e^{iR_j(q,q',t)/\hbar - i\pi m_j(q,q',t)/2}\psi(q'_j,0)\,, \qquad (21.40)$$

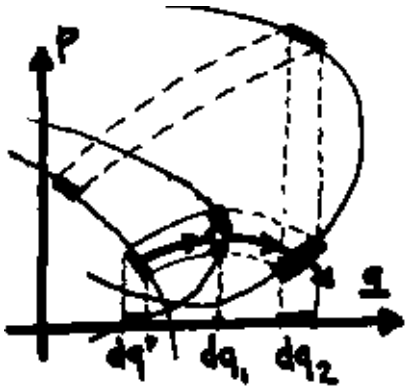


Figure 21.4: Folding of the Lagrangian surface $(q, \partial_q R(q,t))$. The inital surface $(q', p' = \partial_q R(q',0))$ is mapped into the surface $(q(t), p(t))$ some time t later; this surface may develop a fold at $q = q_1$; the volume element dq_1 in the neighborhood of the folding point which steams from some inital volume element dq' is proportional to $\sqrt{dq'}$ instead of dq' at the fold. The Jacobian (21.39), $\partial q'/\partial q$, 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.

each contribution weighted by corresponding density, phase increment and the topological indexphase 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.

21.3 Semiclassical propagator

We saw in sect. 21.1 that the evolution of an initial wave function $\psi(q,0)$ is completely determined by the propagator (21.8). As $K(q,q',t)$ itself satisfies the Schrödinger equation (21.10), 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 (21.11) demands that the propagator at $t = 0$ is a δ-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.

For infinitesimally short times dt 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

$$K(q,q',dt) = A(q,q',dt)e^{\frac{i}{\hbar}R(q,q',dt)}\,.$$

As all particles start at $q = q'$, $R(q,q',dt)$ will be of the form (21.30), that is

$$R(q,q',dt) = p\dot{q}dt - H(q,p)dt\,, \tag{21.41}$$

with $\dot{q} \approx (q-q')/dt$. For Hamiltonians of the form (21.2) we have $\dot{q} = p/m$, which leads to

$$R(q,q',dt) = \frac{m(q-q')^2}{2dt} - V(q)dt\,.$$

Inserting this into our ansatz for the propagator we obtain

$$K_{sc}(q,q',dt) \approx A(q,q',dt)e^{\frac{i}{\hbar}\left(\frac{m}{2dt}(q-q')^2 - V(q)dt\right)}\,. \tag{21.42}$$

For infinitesimal times we can neglecting the term $-V(q)dt$, so $K_{sc}(q,q',dt)$ is a d-dimensional gaussian with width $\sigma^2 = i\hbar dt/m$. This gaussian is a finite width approximation to the Dirac delta function if $A = (m/2\pi i\hbar dt)^{d/2}$, with $A(q,q',dt)$ fixed by the Dirac delta function normalization condition. The correctly normalized propagator for short times dt is therefore

21.8 on p. 508

$$K_{sc}(q,q',dt) \approx \left(\frac{m}{2\pi i\hbar dt}\right)^{d/2} e^{\frac{i}{\hbar}\left(\frac{m(q-q')^2}{2dt} - V(q)dt\right)}\,. \tag{21.43}$$

The short time dynamics of the Lagrangian manifold $(q, \partial_q R)$ which corresponds to the quantum propagator can now be deduced from (21.41); one obtains

$$\frac{\partial R}{\partial q} = p \approx \frac{m}{dt}(q-q')\,,$$

that is the particles start for short times on a Lagrangian manifold which is a plane in phase space, see fig. 21.5. Note, that for $dt \to 0$, 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

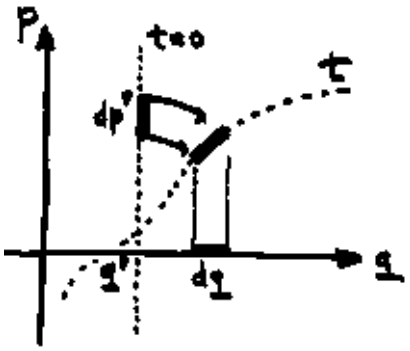


Figure 21.5: Evolution of the semiclassical propagator. The configuration which corresponds to the initial conditions of the propagator is a Lagrangian manifold $q = q'$, that is, a plane parallel to the p axis. The hypothetical particles are thus initially all placed at q' but take on all possible momenta p'. The Jacobian matrix C (21.46) relates an initial volume element in momnetum space dp' to a final configuration space volume dq.

velocities for $t = 0$. The slope of the Lagrangian plane for a short finite time is given as

$$\frac{\partial p_i}{\partial q_j} = -\frac{\partial^2 R}{\partial q_j \partial q'_i} = -\frac{\partial p'_i}{\partial q_j} = \frac{m}{dt}\delta_{ij} \,.$$

The prefactor $(m/dt)^{d/2}$ in (21.43) 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

$$K_{sc}(q, q', dt) = \frac{1}{(2\pi i\hbar)^{d/2}} |\det C|^{1/2} e^{iR(q,q',dt)/\hbar}, \tag{21.44}$$

where

$$C(q, q', dt)_{ij} = \left.\frac{\partial p'_j}{\partial q_i}\right|_{t,q'} = -\frac{\partial^2 R(q, q', dt)}{\partial q_i \partial q'_j} \tag{21.45}$$

The subscript $\cdots|_t$ indicates that the partial derivatives are to be evaluated with t, q' fixed.

The propagator in (21.44) that has been obtained for short times is, however, already more or less in its final form. We only have to evolve our short time approximation of the propagator according to (21.40)

$$K_{sc}(q'', q', t' + dt) = \sum_j |\det \mathbf{j}_j(q'', q, t')|^{1/2} e^{iR_j(q'',q,t')/\hbar - i\pi m_j(q'',q,t')/2} K(q_j, q'_j, dt) \,,$$

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 $m_j = m_j(q'', q', t)$ is the number of singularities in the Jacobian along the trajectory j

from q' to q''. We can write $K_{sc}(q'', q', t' + dt)$ in closed form using the fact that $R(q'', q, t') + R(q, q', dt) = R(q'', q', t' + dt)$ and the multiplicativity of Jacobian determinants, that is

$$\begin{aligned} \det \mathbf{j}(q'', q, t) \det C(q, q', t) &= \det \left.\frac{\partial q}{\partial q''}\right|_t \det \left.\frac{\partial p'}{\partial q}\right|_{q', dt} \\ &= \det \left.\frac{\partial p'}{\partial q''}\right|_{q', t'+dt} = \det C(q'', q', t' + dt) \end{aligned} \tag{21.46}$$

The final form of the semiclassical or *Van Vleck propagator*, is thus

$$K_{sc}(q, q', t) = \sum_j \frac{1}{(2\pi i\hbar)^{d/2}} |\det C_j(q, q', t)|^{1/2} e^{iR_j(q,q',t)/\hbar - im_j\pi/2} \,. \tag{21.47}$$

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 $\sqrt{\det \partial q'/\partial q}$ 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 (21.47) with some care. Unlike the full quantum propagator which satisfies the group property (21.9) exactly the semiclassical propagator performs this only approximately, that is

$$K_{sc}(q, q', t_1 + t_2) \approx \int dq'' \, K_{sc}(q, q'', t_2) K_{sc}(q'', q', t_1) \,. \tag{21.48}$$

The connection can be made explicit by the stationary phase approximation, which we will discuss in sect. 21.4.1. Approximating the integral in (21.48) by integrating only over regions near points q'' at which the phase is stationary, leads to the stationary phase condition

$$\frac{\partial R(q, q'', t_2)}{\partial q''_i} + \frac{\partial R(q'', q, t_1)}{\partial q''_i} = 0. \tag{21.49}$$

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. We will treat the method of stationary phase in more detail in sect. 21.4.1. Unlike the classical evolution of sect. 6.2, 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.

??
on p. ??

21.3.1 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 (21.30) is

$$R(q,q',t) = \frac{m(q-q')^2}{2t}\,. \tag{21.50}$$

The matrix $C(q,q',t)$ from (21.45) can be evaluated explicitly, and the Van Vleck propagator is

$$K_{sc}(q,q',t) = \left(\frac{m}{2\pi i\hbar t}\right)^{d/2} e^{im(q-q')^2/2\hbar t}\,, \tag{21.51}$$

identical to the short times propagator (21.43), 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 (21.20). 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 necessary classical) paths from q' to q, see also remark 21.3.

21.9 on p. 508
21.10 on p. 508
21.11 on p. 508

21.4 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 (21.1). Even though we assumed in the calculation a time independent Hamiltonian of the special form (21.2), the derivation leads to the same final result (21.47) 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 (21.6). It is therefore adventageous to switch from a time representation to an energy representation, that is from the propagator (21.8) to the energy dependent Green's function (21.12). A semiclassical approximation of the Green's function $G_{sc}(q,q',E)$ is given by the Laplace transform (21.12) of the Van Vleck propagator $K_{sc}(q,q',t)$:

$$G_{sc}(q,q',E) = \frac{1}{i\hbar}\int_0^\infty dt\, e^{iEt/\hbar} K_{sc}(q,q',t)\,. \tag{21.52}$$

The expression as it stands is not very useful; in order to evaluate the integral at least approximately we need the method of stationary phase next which we introduce next.

21.4.1 Method of stationary phase

Semiclassical approximations are often based on saddlepoint evaluations of integrals of the type

$$I = \int d^d x\, A(x)\, e^{is\Phi(x)} \tag{21.53}$$

where s is assumed to be a large, real parameter and $\Phi(x)$ is a real-valued function. For large s the phase oscillates rapidly and "averages to zero" everywhere except at the *extremal points* $\Phi'(x_0) = 0$, $\Phi''(x_0) \neq 0$. The method of approximating the integral by its values at extremal points is therefore often called *method of stationary phase*. Restricting ourselves to one-dimensional integrals first, we expand $\Phi(x_0 + \delta x)$ around x_0 to second order in δx, and write

$$I = \int dx\, A(x)\, e^{is(\Phi(x_0)+\frac{1}{2}\Phi''(x_0)\delta x^2+\ldots)}\,.$$

If $A(x)$ varies slowly around x_0 compared to the exponential function we may retain the leading term in an expansion of the amplitude and up to quadratic terms in the phase approximate the integral I by

$$I \approx A(x_0) e^{is\Phi(x_0)} \int_{-\infty}^{\infty} dx\, e^{is\Phi''(x_0)\frac{x^2}{2}} = A(x_0) \left(\frac{2\pi}{s|\Phi''(x_0)|}\right)^{1/2} e^{\frac{\pm i\pi}{4}}\,, \tag{21.54}$$

21.12 on p. 509

where $\pm$ corresponds to positive/negative sign of $\Phi''(x_0)$. The integral (21.54) is known as a *Fresnel integral*. Generalizing this method to d dimensions, consider stationary phase points fulfilling

$$\frac{d}{dx_i}\Phi(x)\bigg|_{x=x_0} = 0 \quad \forall i = 1, \ldots d\,.$$

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 diagonalises D, we can approximate the d-dimensional integral by d one-dimensional Fresnel integrals; the stationary phase estimate of (21.53) is then

$$I \approx \sum_{x_0} (2\pi i/s)^{d/2} \left|\det D(x_0)\right|^{-1/2} A(x_0)\, e^{is\Phi(x_0)-\frac{i\pi}{2}m(x_0)}\,, \tag{21.55}$$

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 $\mathbf{D}(x_0)$.

21.13 on p. 509

The stationary phase approximation is all that is needed for the semiclassical approximation, with the proviso that D in (21.55) has no zero eigenvalues.

21.14 on p. 509

21.15 on p. 509

21.4.2 Long trajectories

When evaluating the integral (21.52) 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 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 no longer valid. The short time contributions therefore have to be treated separately, which we will do in sect. 21.4.3.

The stationary phase points t^* of the integrand in (21.52) are given by the condition

$$\frac{\partial}{\partial t} R(q,q',t^*) + E = 0\,. \tag{21.56}$$

We recognize this condition as the solution of (21.35), 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,

$$R(q,q',t) + Et = R(q,q',t^*) + \frac{1}{2}(t-t^*)^2 \frac{\partial^2}{\partial t^2} R(q,q',t^*) + \cdots$$

the stationary phase approximation of the integral corresponding to a specific branch j of the Van Vleck propagator (21.47) yields

$$G_j(q,q',E) = \frac{1}{i\hbar(2i\pi\hbar)^{(d-1)/2}} \left|\det C_j \left(\frac{\partial^2 R_j}{\partial t^2}\right)^{-1}\right|^{1/2} e^{\frac{i}{\hbar}S_j - \frac{i\pi}{2}m_j}\,, \tag{21.57}$$

where $m_j = m_j(q, q', E)$ now includes a possible additional phase arising from the time stationary phase integration (21.54), 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 (21.33)

$$S(q, q', E) = R(q, q', t^*) + Et^* , \qquad \text{with} \quad t^* = t^*(q, q', E) , \tag{21.58}$$

the Legendre transform of the Hamilton's principal function. Note that the partial derivative of the action (21.58) 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

$$\frac{\partial S(q, q', E)}{\partial q_i} = \frac{\partial R(q, q', t^*)}{\partial q_i} , \tag{21.59}$$

due to the stationary phase condition (21.56), so the definition of momentum as a partial derivative with respect to q remains unaltered by the Legendre transform from time to energy domain.

21.17 on p. 510

Next we will simplify the amplitude term in (21.57) and rewrite it as an explicit function of the energy. Consider the $[(d+1)\times(d+1)]$ matrix

$$D(q, q', E) = \begin{pmatrix} \frac{\partial^2 S}{\partial q' \partial q} & \frac{\partial^2 S}{\partial q' \partial E} \\ \frac{\partial^2 S}{\partial q \partial E} & \frac{\partial^2 S}{\partial E^2} \end{pmatrix} = \begin{pmatrix} -\frac{\partial p'}{\partial q} & -\frac{\partial p'}{\partial E} \\ \frac{\partial t}{\partial q} & \frac{\partial t}{\partial E} \end{pmatrix} , \tag{21.60}$$

where $S = S(q, q', E)$ and we used (21.31–21.34) here to obtain the left hand side of (21.60). The minus signs follow from observing from the definition of (21.32) 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

$$\begin{aligned} \det D &= (-1)^{d+1} \left\| \frac{\partial(p', t)}{\partial(q, E)} \right\|_{q'} = (-1)^{d+1} \left\| \frac{\partial(p', t)}{\partial(q, t)} \frac{\partial(q, t)}{\partial(q, E)} \right\|_{q'} \\ &= (-1)^{d+1} \left\| \frac{\partial p'}{\partial q} \right\|_{t,q'} \left\| \frac{\partial t}{\partial E} \right\|_{q',q} = -\det C \left(\frac{\partial^2 R}{\partial t^2} \right)^{-1} . \end{aligned}$$

21.18 on p. 510

We use here the notation $\|.\|_{q',t}$ for a Jacobian determinant with partial deriva-

tives evaluated at t, q' fixed, and likewise for other subscripts. Using the relation (21.36) 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 (21.45) and the amplitude factor arising from the stationary phase approximation. The amplitude in (21.57) 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 (21.45).

We can now write down the semiclassical approximation of the contribution of the jth trajectory to the Green's function (21.57) in explicitly energy dependent form:

$$G_j(q,q',E) = \frac{1}{i\hbar(2i\pi\hbar)^{(d-1)/2}} \left|\det D^j\right|^{1/2} e^{\frac{i}{\hbar}S_j - \frac{i\pi}{2}m_j} \,. \tag{21.61}$$

However, for no purposes 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 partial differential equation for $S(q,q',E)$, that is

$$H(q, \frac{\partial S}{\partial q}) = E \,,$$

one obtains

$$\begin{aligned} \frac{\partial}{\partial q'_i} H(q,p) &= 0 = \frac{\partial H}{\partial p_j}\frac{\partial p_j}{\partial q'_i} = \dot{q}_j \frac{\partial^2 S}{\partial q_j \partial q'_i} \\ \frac{\partial}{\partial q_i} H(q',p') &= 0 = \frac{\partial^2 S}{\partial q_i \partial q'_j}\dot{q}'_j \,, \end{aligned} \tag{21.62}$$

that is the sub-matrix $\partial^2 S/\partial q_i \partial q'_j$ has (left- and right-) eigenvectors corresponding to an eigenvalue 0. In the local coordinate system

$$\mathbf{q} = (q_\parallel, q_{\perp 1}, q_{\perp 2}, \cdots, q_{\perp(d-1)}) \,, \qquad \text{with} \quad \dot{\mathbf{q}} = (\dot{q}, 0, 0, \cdots, 0)$$

in which the longitudinal coordinate axis $q_\parallel$ points along the velocity vector $\dot{q}$, the matrix of variations of $S(q,q',E)$ has a column and a row of zeros as (21.62) takes form

$$\dot{q}\frac{\partial^2 S}{\partial q_\parallel \partial q'_i} = \frac{\partial^2 S}{\partial q_i \partial q'_\parallel}\dot{q}' = 0 \,.$$

The initial and final velocities are non-vanishing except for points $|\dot{q}| = 0$. These are the turning points (where all energy is potential), and we assume that neither q nor q' is a turning point (in our application - periodic orbits - we can always chose $q = q'$ not a truning point). In the local coordinate system with one axis along the trajectory and all other perpendicular to it the determinant of (21.60) is of the form

$$\det D(q,q',E) = (-1)^{d+1} \begin{Vmatrix} 0 & 0 & \frac{\partial^2 S}{\partial E \partial q'_\parallel} \\ 0 & \frac{\partial^2\ S}{\partial q_\perp \partial q'_\perp} & * \\ \frac{\partial^2 S}{\partial q_\parallel \partial E} & * & * \end{Vmatrix} . \tag{21.63}$$

The corner entries can be evaluated using (21.34)

$$\frac{\partial^2 S}{\partial q_\parallel \partial E} = \frac{\partial}{\partial q_\parallel} t = \frac{1}{\dot{q}} , \qquad \frac{\partial^2 S}{\partial E \partial q'_\parallel} = \frac{1}{\dot{q}'} .$$

As the $q_\parallel$ axis points along the velocity direction, velocities $\dot{q}$, $\dot{q}'$ are by construction almost always positive non-vanishing numbers. In this way the determinant of the $[(d+1)\times(d+1)]$ dimensional matrix $D(q,q',E)$ can essentially be reduced to the determinant of a $[(d-1)\times(d-1)]$ dimensional *transverse* matrix $D_\perp(q,q',E)$

$$\begin{aligned} \det D(q,q',E) &= \frac{1}{\dot{q}\dot{q}'} \det D_\perp(q,q',E) \\ D_\perp(q,q',E)_{ik} &= -\frac{\partial^2 S(q,q',E)}{\partial q_{\perp i} \partial q'_{\perp k}} . \end{aligned} \tag{21.64}$$

21.19 on p. 511

Putting everything together we obtain the jth trajectory contribution to the semiclassical Green's function

$$G_j(q,q',E) = \frac{1}{i\hbar(2\pi\hbar)^{(d-1)/2}} \frac{1}{|\dot{q}\dot{q}'|^{1/2}} \left|\det D_\perp^j\right|^{1/2} e^{\frac{i}{\hbar}S_j - \frac{i\pi}{2}m_j} , \tag{21.65}$$

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 velocities $\dot{q}$, $\dot{q}'$ also depend on (q,q',E) and the trajectory j. While in the case of the propagator the initial momentum variations $\delta p'$ are unrestricted, for the Green's function the $(\delta q', \delta p')$ variations are restricted to the constant energy shell; the appearance of the $1/\dot{q}\dot{q}'$ weights in the Green's function can be traced to this constraint.

21.4.3 Short trajectories

The stationary phase method cannot be used when t^* is small, both because we cannot extend the integration in (21.54) 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 (21.43)

$$G_0(q,q',E) = \frac{1}{i\hbar}\int_0^\infty dt \left(\frac{m}{2\pi i\hbar t}\right)^{d/2} e^{\frac{i}{\hbar}(\frac{m(q-q')^2}{2t} - V(q)t + Et)}.$$

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

$$G_0(q,q',E) \approx -\frac{im}{2\hbar^2}\left(\frac{\sqrt{2m(E-V)}}{2\pi\hbar|q-q'|}\right)^{\frac{d-2}{2}} H^+_{\frac{d-2}{2}}(S_0(q,q',E)/\hbar). \qquad (21.66)$$

There is nothing sacred about the Hankel function - it is merely a useful observation, as for special functions the short wavelength asymptotics comes for free and can be found in standard reference books. The short distance Green's function approximation is valid when $S_0(q,q',E) \leq \hbar$.

Commentary

Remark **21.1** Limit $\hbar \to 0$. The semiclassical limit "$\hbar \to 0$" discussed in sect. 21.2 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 [6] is $1.054571596(82)\ 10^{-34}$ Js.

Remark **21.2** Madelung's fluid dynamics. A very different interpretation of (21.21–21.22) has been given by Madelung [2], and then built upon by Bohm [4] and others [32, 5]. Keeping the $\hbar$ dependent term in (21.21), the ordinary differential equations driving the flow (21.28) have to be altered; if the Hamiltonian can be written as kinetic plus potential term $V(q)$ as in (21.2), the $\hbar^2$ term modifies the p equation of motion as

$$\dot{p_i} = -\frac{\partial}{\partial q_i}\left(V(q) + Q(q,t)\right) , \tag{21.67}$$

where, for the example at hand,

$$Q(q,t) = -\frac{\hbar^2}{2m}\frac{1}{\sqrt{\rho}}\frac{\partial^2}{\partial q^2}\sqrt{\rho} \tag{21.68}$$

interpreted by Bohm [4] as the "quantum potential". Madelung observed that Hamilton's equation for the momentum (21.67) can be rewritten as

$$\frac{\partial v_i}{\partial t} + \left(v\cdot\frac{\partial}{\partial q}\right)v_i = -\frac{1}{m}\frac{\partial V}{\partial q_i} - \frac{1}{m\rho}\frac{\partial}{\partial q_j}\sigma_{ij} , \tag{21.69}$$

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 [32] in sect. 21.2.3. 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 ρ dependent stresses. The "hydrodynamic" interpretation of quantum mechanics has, however, not been very fruitful in practice.

Remark **21.3** Path integrals. The semiclassical propagator (21.47) 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 (21.51) 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 [22] added the topolagical indices when deriving the semiclassical propagator from Feynman's path integral by stationary phase conditions.

Résumé

The aim of the semiclassical or short-wavelength methods is to approximate a solution of the Schrödinger equation with a semi-classical 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*

$$K_{sc}(q,q',t) = \frac{1}{(2\pi i\hbar)^{d/2}} \sum_j \left\|\partial p'/\partial q\right\|_j^{1/2} e^{\frac{i}{\hbar}R_j - \frac{i\pi}{2}m_j} ,$$

where the topological index $m_j(q,q',t)$ counts the number of the direction reversal along the jth 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 enformtion is extracted by means of a Laplace transform of the propagator which yields the energy dependent *semiclassical Green's function*

$$G_{sc}(q,q',E) = G_0(q,q',E) + \frac{1}{i\hbar(2\pi i\hbar)^{\frac{(d-1)}{2}}} \sum_j \frac{\left\|\partial p'_\perp/\partial q_\perp\right\|_j^{1/2}}{\left|\dot{q}\dot{q}'\right|_j^{1/2}} e^{i\frac{R_j}{\hbar} - im_j\frac{\pi}{2}} \quad (21.70)$$

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 (21.65) going from q' to q with fixed energy E, with $S_j(q,q',E) \gg \hbar$.

References

[21.1] 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*, Vol. **6**: *The Berlin Years: Writings, 1914-1917*, A. Engel, transl. and E. Schucking, ed. (Princeton University Press, Princeton, New Jersey 1997).

[21.2] E. Madelung, *Zeitschr. fr Physik* **40**, 332 (1926).

[21.3] E. Schrödinger, *Annalen der Physik* **79**, 361, 489; **80**, 437; **81**, 109 (1926).

[21.4] D. Bohm, *Phys. Rev.* **85**, 166 (1952).

[21.5] 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).

[21.6] `http://physics.nist.gov/cgi-bin/cuu`

Exercises

21.1 Dirac delta function, Lorentzian representation. Derive Eq. (21.15).

21.2 WKB ansatz. Try to show that no other ansatz gives a meaningful definition of the momentum in the $\hbar \to 0$ limit.

21.3 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 (21.21) and (21.22) and solve them.

21.4 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 (21.21) and (21.22).

21.5 d-dimensional quadratic potentials. Generalize the above method to general d-dimensional quadratic potentials.

21.6 Free particle 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.

21.7 Time evolution of R. Calculate the time evolution of $R(q,0) = a + bq + cq^2$ for a 1-dimensional harmonic oscillator using (21.30) and (21.31).

21.8 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 (**??**).

21.9 d-dimensional free particle propagator. Verify the results in sect. 21.3.1; show explicitely that (21.51), the semiclassical Van Vleck propagator in d dimensions, solves the Schrödinger's equation.

21.10 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.

21.11 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.

21.12 Fresnel integral. Show that

$$\int_{-\infty}^{\infty} dx\, e^{iax^2/2} = \left(\frac{2\pi}{|a|}\right)^{1/2} e^{\frac{i\pi m}{4}} \tag{21.71}$$

where $m = a/|a|$ is the sign of a.

21.13 d-dimensional gaussian integrals. Show that the gaussian integral in d-dimensions is given by

$$\frac{1}{(2\pi)^{d/2}} \int d^d x e^{-\frac{1}{2}x^T\cdot M^{-1}\cdot x + x\cdot J} \;=\; |\det M|^{\frac{1}{2}}\, e^{\frac{1}{2}J^T\cdot M\cdot J}\,, \tag{21.72}$$

where M is a real positive definite $[d\times d]$ matrix, that is a matrix with strictly positive eigenvalues. x, J are d-dimensional vectors, on x^T is the transpose of x.

21.14 Stationary phase approximation. All semiclassical approximations are based on saddlepoint evaluations of integrals of type

$$I = \int d^d x A(x) e^{i\Phi(x)/\hbar} \tag{21.73}$$

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.

21.15 Sterling formula for $n!$. Compute an approximate value of $n!$ for large n with the help of stationary phase approximation. Hint: $n! = \int_0^\infty dt\, t^{n-1}e^{-t}$.

21.16 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). In such calculations one meets Airy functions integrals of the form

$$Ai(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} dy\, e^{i(xy-\frac{y^3}{3})}\,. \tag{21.74}$$

Calculate the Airy function $Ai(x)$ by stationary phase approximation. What happens when considering the limit $x \to 0$. Estimate for which value of x the stationary phase approximation breaks down. Give it a go.

21.17 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.

21.18 A usefull determinant identity. Show that the following two determinants equal each other:
$A[(n+1)\times(n{+}?)]determinant$

$$\det\left(M'_n\right) \;=\; \overset{n+1}{\begin{bmatrix} x_{1,1} & \dots & x_{1,n} & y_1 \\ \vdots & \ddots & \vdots & \vdots \\ x_{n,1} & \dots & x_{n,n} & y_n \\ z_1 & \dots & z_n & E \end{bmatrix}}\; n+1 \tag{21.75}$$

and the $[n\times n]$ determinant:

$$E\det\left(M_n\right) = E\;\overset{n}{\begin{bmatrix} x_{1,1}-y_1z_1E^{-1} & \dots & x_{1,n}-y_1z_nE^{-1} \\ \vdots & \ddots & \vdots \\ x_{n,1}-y_nz_1E^{-1} & \dots & x_{n,n}-y_nz_nE^{-1} \end{bmatrix}}\; n \tag{21.76}$$

21.19 Free particle semiclassical Green's functions. Calculate the semiclassical Green's functions for the systems of exercise 21.17.

Chapter 22

Semiclassical quantization

(G. Vattay, G. Tanner and P. Cvitanović)

> If there exist fewer than ℓ integrals of type (14), as is the case, for example, according to POINCARÉ in the three-body problem, then the p_i are not expressible by the q_i and the quantum condition of SOMMERFELD-EPSTEIN fails also in the slightly generalized form that has been given here.
>
> A. Einstein

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 **??** 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 24.

22.1 Trace formula

Our next task is to evaluate the Green's function trace (21.13) of 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 which start and end in q after finite time, and the "zero length" trajectories whose lengths approach zero as $q' \to q$.

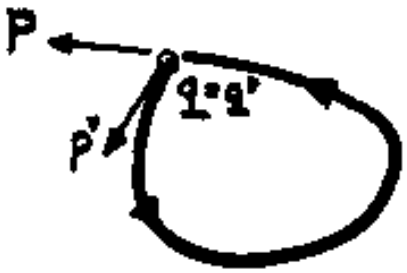


Figure 22.1: A returning trajectory in configuration space. The orbit is periodic in the full phase space only if the initial and the final momenta of a returning trajectory coincide as well.

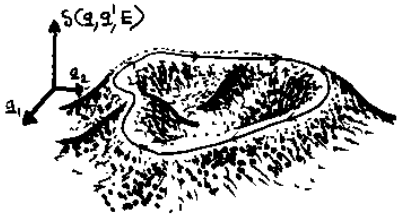


Figure 22.2: A romanticized sketch of $S_p(E) = S(q,q,E) = \oint p(q,E)dq$ landscape orbit. Unstable periodic orbits traverse isolated ridges and saddles of the mountainous landscape of the action $S(q_\parallel, q_\perp, E)$. Along a periodic orbit $S_p(E)$ is constant; in the transverse directions it generically changes quadratically.

First we work out the contributions coming from the finite time *returning* classical orbits, that is 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,

$$\left.\frac{\partial S_j(q,q',E)}{\partial q_i}\right|_{q'=q} + \left.\frac{\partial S_j(q,q',E)}{\partial q'_i}\right|_{q'=q} = 0\,,$$

meaning that the initial and final momenta (21.59) of contributing trajectories should coincide

$$p_j(q,q,E) - p'_j(q,q,E) = 0\,, \qquad q \in j\text{th periodic orbit}\,, \tag{22.1}$$

so the trace receives contributions only from those long classical trajectories which are *periodic* in the full phase space.

For a returning 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 jth periodic orbit contribution to the trace of the semiclassical Green's function in the intrinsic coordinates is

$$\mathrm{tr}\, G_j(E) = \frac{1}{i\hbar(2\pi\hbar)^{(d-1)/2}} \oint_j \frac{dq_\parallel}{\dot{q}} \int_j d^{d-1}q_\perp |\det D_\perp^j|^{1/2} e^{\frac{i}{\hbar}S_j - \frac{i\pi}{2}m_j}\,,$$

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 $m(q_\parallel, q_\perp, E)$ is an integer which generically does not change in the infinitesimal neighborhood of an isolated periodic orbit.

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.

The ratio $D_{\perp j}/D'_{\perp j}$ 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

$$\begin{aligned} \frac{\det D'_\perp}{\det D_\perp} &= \left\| \frac{\partial (p_\perp - p'_\perp, q_\perp - q'_\perp)}{\partial (q'_\perp, p'_\perp)} \right\| = \left\| \frac{\partial (x_\perp - x'_\perp)}{\partial x'_\perp} \right\| \\ &= \det\, (\mathbf{J} - \mathbf{1}) \,, \end{aligned} \tag{22.2}$$

in terms of the monodromy matrix $\mathbf{J}$ for a surface of section transverse to the orbit within the constant energy $E = H(q, p)$ shell.

The classical periodic orbit action $S_j(E) = \oint p(q_\parallel, E) dq_\parallel$ does not depend on the starting point $q_\parallel$ along the orbit, see fig. 22.2. The eigenvalues of the monodromy matrix are independent of where $\mathbf{J}_j$ is evaluated along the orbit, so $\det(1 - \mathbf{J}_j)$ can also be taken out of the the $q_\parallel$ integral

$$\operatorname{tr} G_j(E) = \frac{1}{i\hbar} \sum_j \frac{1}{|\det(1 - \mathbf{J}_j)|^{1/2}} e^{r(\frac{i}{\hbar} S_j - \frac{i\pi}{2} m_j)} \oint \frac{dq_\parallel}{\dot{q}_\parallel} .$$

Here we have assumed that $\mathbf{J}_j$ has no marginal eigenvalues. The determinant of the monodromy matrix, the action $S_p(E) = \oint p(q_\parallel, E) dq_\parallel$ and the topological index are all classical invariants of the periodic orbit, independent of $q_\parallel$. 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 rth repeat they simply get multiplied by r. The monodromy matrix of the rth repeat of a prime cycle p is (by the chain rule for derivatives) $\mathbf{J}_p^r$, where $\mathbf{J}_p$ is the prime cycle monodromy matrix. Let us denote the time period of the prime cycle p, the single, shortest traversal of a periodic orbit by T_p. The remaining integral can be carried out by change of variables $dt = dq_\parallel / \dot{q}(t)$

$$\int_0^{L_j} \frac{dq_\parallel}{\dot{q}(t)} = \int_0^{T_p} dt = T_p,$$

as 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 (7.19) - that derivation takes only three pages of text. Regrettably, in the quantum case we do not know of an hones derivation that takes less 30 pages. The final result, the *Gutzwiller trace formula*

$$\operatorname{tr} G_{sc}(E) = \operatorname{tr} G_0(E) + \frac{1}{i\hbar} \sum_p T_p \sum_{r=1}^{\infty} \frac{1}{|\det(1 - \mathbf{J}_p^r)|^{1/2}} e^{r(\frac{i}{\hbar} S_p - \frac{i\pi}{2} m_p)} , \quad (22.3)$$

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 2-d 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.

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.

22.1.1 Semiclassical density of states

The semiclassical contribution to the density of states (21.13) is given by the imaginary part of the Gutzwiller trace formula (22.3) multiplied with $-1/\pi$. The contribution coming from the zero length trajectories is the imaginary part of (21.66) for $q' \to q$ integrated over the configuration space

$$d_0(E) = -\frac{1}{\pi} \int d^d q \, \mathrm{Im}\, G_0(q, q, E),$$

$$\mathrm{Im}\, G_0(q, q, E) = -\lim_{q' \to q} \frac{m}{2\hbar^2} \left(\frac{\sqrt{2m(E-V)}}{2\pi\hbar |q - q'|} \right)^{\frac{d-2}{2}} J^+_{\frac{d-2}{2}}(S_0(q, q', E)/\hbar).$$

where $G_0(q, q', E)$ is the short time Green's function (21.66). Using the Bessel function $J^+_\nu = \mathrm{Im}\, H^+_\nu$ asymptotic estimate

$$J_\nu(z) \approx \frac{1}{\Gamma(\nu + 1)} \left(\frac{z}{2} \right)^\nu \quad \text{for } |z| << 1.$$

we obtain the zero length trajectories contribution to the density of states

$$d_0(E) = \frac{m}{\hbar^d 2^{d-1} \pi^{d/2} \Gamma(d/2)} \int_{V(q)<E} d^d q \, [2m(E - V(q))]^{(d-2)/2} \,. \tag{22.4}$$

The result is the same as (**??**), the average density estimate $\bar{d}(E)$. The initial wild guess is indeed correct, and 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, $d_{sc}(E) = \bar{d}(E) + d_{osc}(E)$, where

$$d_{osc}(E) = \frac{1}{\pi\hbar} \sum_p T_p \sum_{r=1}^{\infty} \frac{\cos(r S_p(E)/\hbar - r m_p \pi/2)}{|\det\,(1 - \mathbf{J}^r_p)|^{1/2}} \tag{22.5}$$

follows from the trace formula (22.3).

22.1.2 Regularization of the trace

The real part of the $q' \to q$ zero length Green's function (21.66) is ultraviolet divergent in dimensions $d > 1$, and so is its formal trace (21.13). The short distance behavior of the real part of the Green's function can be extracted from the real part of (21.66) 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

$$\mathrm{tr}\, G_{reg}(E) = -i\pi\bar{d}(E) + \frac{1}{i\hbar}\sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r(\frac{i}{\hbar}S_p(E) - \frac{i\pi}{2}m_p(E))}}{|\det(1-\mathbf{J}_p^r)|^{1/2}} . \tag{22.6}$$

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.

22.2 Semiclassical spectral determinant

> The theoreticians now work with an audacity unheard of in earlier times, at present no physical law is considered assured beyond doubt, each and every physical truth is open to dispute. It often looks as if the time of chaos again is drawing near in theoretical physics.
> M. Planck, *Phys. Zeitschr.* **11**, 922 (1910)

The problem with trace formulas is that they diverge where we need them, at

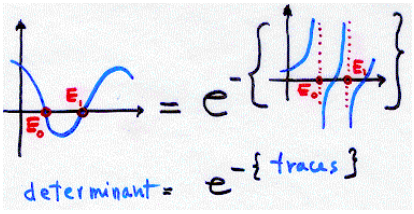


Figure 22.3: A sketch of how spectral determinants convert poles into zeros: The trace shows $1/(E-E_n)$ type singularities at the eigenenergies while the spectral determinant goes smoothly through zeroes.

the individual energy eigenvalues. What to do? Much of the quantum chaology literature responds to the challenge of wrestling the trace formulas by replacing the delta functions in the density of states (21.14) 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 characteristic determinants.

A better way to compute energy levels is to construct the spectral determinant whose zeroes yield the eigenenergies, $\det{}_{sc}(H-E_n)=0$. A first guess might be that the spectral determinant is the Hadamard product of form

$$f(E)=\prod_n (E-E_n),$$

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 22.1.2. Here we offer an impressionistic sketch of regularization.

The logarithmic derivative of $f(E)$ is the formal trace of the Green's function

$$-\frac{d}{dE}\ln f(E)=\sum_n \frac{1}{E-E_n}=\mathrm{tr}\, G(E).$$

This quantity, not surprisingly, is divergent again. This relation however opens a way to derive a convergent version of $f(E)$, by replacing the trace with the regularized trace

$$-\frac{d}{dE}\ln\det{}_{sc}(\hat{(}H)-E)(E)=\mathrm{tr}\, G_{reg}(E).$$

The regularized trace still has $1/(E-E_n)$ poles at the semiclassical eigenenergies, poles which can be generated only if $\det{}_{sc}(\hat{(}H)-E)(E)$ has a zero at $E=E_n$, see fig. 22.3. By integrating and exponentiating we obtain

$$\det{}_{sc}(\hat{(}H)-E)(E)=\exp\left(-\int^E dE'\,\mathrm{tr}\, G_{reg}(E')\right)$$

Now we can use (22.6) and integrate the terms coming from periodic orbits, using the relation (21.34) between the action and the period of a periodic orbit, $dS_p(E) = T_p(E)dE$, and the relation (21.17) between the density of states and the spectral staircase, $d\bar{N}(E) = \bar{d}(E)dE$. We obtain the *semiclassical zeta function*

$$\det{}_{sc}(\hat{(}H) - E)(E) = e^{i\pi\bar{N}(E)} \exp\left(- \sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{ir(S_p/\hbar - m_p\pi/2)}}{|\det\left(1 - \mathbf{J}_p^r\right)|^{1/2}} \right) . \quad (22.7)$$

We already know from the study of classical evolution operator spectra of chapter 8 that this can be evaluated by means of cycle expons of chapter 13. As always, the cycle action S_p, the topological index m_p and the monodromy matrix $\mathbf{J}_p$ eigenvalues are cycle invariants that depend only on the energy E.

22.3 One-dimensional 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 configuration dimension.

In one dimension the semiclassical density of states follows from the one-dimensional form of the oscillating density (22.5) and of the average density (**??**)

$$d(E) = \frac{T_p(E)}{2\pi\hbar} + \sum_r \frac{T_p(E)}{\pi\hbar} \cos(rS_p(E)/\hbar - rm_p(E)\pi/2). \quad (22.8)$$

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 (22.8) can be rewritten by using the Fourier series expansion of a delta spike train

$$\sum_{n=-\infty}^{\infty} \delta(x-n) = \sum_{k=-\infty}^{\infty} e^{i2\pi kx} = 1 + \sum_{k=1}^{\infty} 2\cos(2\pi kx).$$

We obtain

$$d(E) = \frac{T_p(E)}{2\pi\hbar} \sum_n \delta(S_p(E)/2\pi\hbar - m_p(E)/4 - n). \quad (22.9)$$

This expression can be simplified by using the relation (21.34) between T_p and S_p, and the identity (5.7) $\delta(x-x^*) = |f'(x)|\delta(f(x))$, 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 (22.9)

$$S_p(E_n)/2\pi\hbar = n + m_p(E)/4.$$

These are precisely the *Bohr-Sommerfeld quantized energies* E_n satisfying the condition

$$S_p(E_n)/2\pi = \frac{1}{2\pi}\oint p(q, E_n)dq = \hbar\left(n + \frac{m_p(E)}{4}\right). \tag{22.10}$$

In this way the trace formula recovers the well known 1-dimensional 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

$$\bar{N}(E) = n - \frac{1}{2} = S_p(E)/2\pi\hbar - m_p(E)/4 - \frac{1}{2}. \tag{22.11}$$

The 1-dimensional spectral determinant follows from (22.7) by dropping the monodromy matrix part and using (22.11)

$$\det{}_{sc}(\hat{(}H) - E)(E) = e^{-\frac{i}{2\hbar}S_p + \frac{i\pi}{2}m_p} e^{-\sum_r \frac{1}{r} e^{\frac{i}{\hbar}rS_p - \frac{i\pi}{2}rm_p}}. \tag{22.12}$$

Summation yields a logarithm by $\sum_r t^r/r = -\ln(1-t)$ and we get

$$\begin{aligned}\det{}_{sc}(\hat{(}H) - E)(E) &= e^{-\frac{i}{2\hbar}S_p + \frac{im_p}{4} + \frac{i\pi}{2}}(1 - e^{\frac{i}{\hbar}S_p - i\frac{m_p}{2}}) \\ &= 2\sin\left(S_p(E)/\hbar - m_p(E)/4\right).\end{aligned}$$

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 (22.10).

22.4 Two-dimensional systems

For flows in two configuration dimensions the monodromy matrix $\mathbf{J}_p$ has two eigenvalues Λ_p and $1/\Lambda_p$, as explained in sect. 4.4. Isolated periodic orbits can be elliptic or hyperbolic. Here we discuss only the hyperbolic case, when the eigenvalues are real and their absolute value is not equal to one. The determinant appearing in the trace formulas can be written in terms of the expanding eigenvalue as

$$|\det\left(1-\mathbf{J}_p^r\right)|^{1/2} = |\Lambda_p^r|^{1/2}\left(1-1/\Lambda_p^r\right)\ ,$$

and its inverse can be expanded as a geometric series

$$\frac{1}{|\det\left(1-\mathbf{J}_p^r\right)|^{1/2}} = \sum_m \frac{1}{|\Lambda_p^r|^{1/2}\Lambda_p^{mr}}.$$

With the 2-dimensional expression for the average density of states (**??**) the spectral determinant becomes

$$\begin{aligned}\det{}_{sc}(\hat{(}H)-E)(E) &= e^{i\frac{m\mathcal{A}E}{2\hbar^2}}\exp\left(-\sum_p\sum_{r=1}^{\infty}\sum_{k=0}^{\infty}\frac{e^{irS_p/\hbar-irm_p\pi/2}}{r|\Lambda_p^r|^{1/2}\Lambda_p^{kr}}\right)\\ &= e^{i\frac{m\mathcal{A}E}{2\hbar^2}}\prod_p\prod_{k=0}^{\infty}\left(1-\frac{1}{|\Lambda_p|^{1/2}\Lambda_p^k}e^{\frac{i}{\hbar}S_p-\frac{i\pi}{2}m_p}\right)\ . \qquad (22.13)\end{aligned}$$

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 semiconductors, 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. In classical dynamics trace formulas (**??**) hide under a variety of appelations such as the $f-\alpha$ or multifractal formalism; in quantum mechanics they are known as the Gutzwiller trace formulas. The trace of the semiclassical Green's function

$$\operatorname{tr} G_{sc}(E) = \int d^d q\, G_{sc}(q,q,E).$$

is given by a sum over the periodic orbits (22.6).

While easiest to derive, in calculations the trace formulas are hard to use for anything other than the leading eigenvalue estimates, as they tend to be divergent in the region of physical interest.

Zeros of *Ruelle or dynamical zeta functions*

$$1/\zeta(s) = \prod_p (1-t_p), \quad t_p = \frac{1}{\sqrt{\Lambda_p}} e^{\frac{i}{\hbar}S_p - i\pi m_p/2}$$

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.

Selberg-type zeta functions, Fredholm determinants, spectral 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-dimensional Selberg-type zeta function (22.13)

$$\det{}_{sc}(\hat{(}H) - E)(E) = \prod_p \prod_{k=0}^{\infty} \left(1 - \frac{e^{iS_p/\hbar - i\pi m_p/2}}{|\Lambda_p|^{1/2}\Lambda_p^k}\right)$$

is a typical example. Most periodic orbit calculations are based on cycle expansions of such determinants.

As we have assumed 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 are valid only for the hyperbolic and elliptic periodic orbits.

For the deterministic dynamical flows and number theory, 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?

References

[22.1] M. Jammer, *The Conceptual Development of Quantum mechanics* (McGraw-Hill, New York 1966).

[22.2] J. Mehra and H. Rechtenberg, *The Historical Development of the Quantum Theory* (Springer, New York 1982).

[22.3] J.W.S. Rayleigh, *The Theory of Sound* (Macmillan, London 1896; reprinted by Dover, New York 1945).

[22.4] Young, (180?).

[22.5] Airy, (1838).

[22.6] Stokes, (1847).

[22.7] Gouy, (1899).

[22.8] Jefferies, (1920).

[22.9] J.B. Keller, *Ann. Phys. (N.Y.)* **4**, 180 (1958).

[22.10] J.B. Keller and S.I. Rubinow, *Ann. Phys. (N.Y.)* **9**, 24 (1960).

[22.11] 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).

[22.12] J.B. Keller, *Calculus of Variations***27**, (1958).

[22.13] V.P. Maslov, *Théorie des Perturbations et Métodes Asymptotiques* (Dunod, Paris, 1972).

[22.14] V.P. Maslov and M.V. Fedoriuk, *Semi-Classical Approximation in Quantum Mechanics* (Reidel, Boston 1981).

[22.15] V.I. Arnold, *Functional Anal. Appl.* **1**, 1 (1967).

[22.16] J.E. Marsden and T.S. Ratiu, *Introduction to Mechanics and Symmetry* (Springer-Verlag, New York, 1994).

[22.17] N. Bleistein and R.A. Handelsman, *Asymptotic Expansions of Integrals* (Dover, New York 1986).

[22.18] R.G. Littlejohn, J. Stat. Phys. **68**, 7 (1992).

[22.19] L.D. Landau and E.M. Lifshitz, *Mechanics* (Pergamon, London, 1959).

[22.20] R.G. Littlejohn, *"Semiclassical structure of trace formulas"*, in G. Casati and B. Chirikov, eds., *Quantum Chaos*, (Cambridge University Press, Cambridge 1994).

[22.21] V.I. Arnold, *Ordinary Differential Equations* (MIT Press, Cambridge, Mass. 1978).

[22.22] M.C. Gutzwiller, *J. Math. Phys.* **8**, 1979 (1967); **10**, 1004 (1969); **11**, 1791 (1970); **12**, 343 (1971).

[22.23] M.C. Gutzwiller, *J. Math. Phys.* **12**, 343 (1971)

[22.24] M.C. Gutzwiller, *Physica* **D5**, 183 (1982)

[22.25] M.C. Gutzwiller, *J. Phys. Chem.* **92**, 3154 (1984).

[22.26] A. Voros, *J. Phys.* **A 21**, 685 (1988).

[22.27] 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).

[22.28] A. Wirzba, *CHAOS* **2**, 77 (1992).

[22.29] 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`.

[22.30] E.B. Bogomolny, CHAOS **2**, 5 (1992).

[22.31] E.B. Bogomolny, *Nonlinearity* **5**, 805 (1992).

[22.32] E. Schrödinger, *Annalen der Physik* **79**, 361, 489; **80**, 437, **81**, 109 (1926).

[22.33] E.A. Spiegel

[22.34] M. Kline, *Mathematical Thought from Ancient to Modern Times* (Oxford Univ. Press, Oxford 1972); on Monge and theory of characteristics - chapter 22.7.

[22.35] E.T. Bell, *Men of Mathematics* (Penguin, London 1937).

[22.36] R.P. Feynman, *Statistical Physics* (Addison Wesley, New York 1990).

[22.37] Goldstein, *Mechanics*, chapter 9.

[22.38] G. Tanner and D. Wintgen, *CHAOS* **2**, 53 (1992).

[22.39] P. Cvitanović and F. Christiansen, *CHAOS* **2**, 61 (1992).

[22.40] M.V. Berry and J.P. Keating, *J. Phys.* **A 23**, 4839 (1990).

[22.41] H.H. Rugh, *"Generalized Fredholm determinants and Selberg zeta functions for Axiom A dynamical systems"*, *Ergodic Theory Dynamical Systems* **16**, 805 (1996).

[22.42] B. Eckhard and G. Russberg, Phys. Rev. **E 47**, 1578 (1993).

[22.43] D. Ruelle, *Statistical Mechanics, Thermodynamical Formalism* (Addison-Wesley, Reading MA, 1987).

[22.44] P. Szépfalusy, T. Tél, A. Csordás and Z. Kovács, Phys. Rev. **A 36**, 3525 (1987).

[22.45] H.H. Rugh, Nonlinearity **5**, 1237 (1992) and H.H. Rugh, *Ph.D. Thesis* (Niels Bohr Institute, 1993).

[22.46] P. Cvitanović, P.E. Rosenqvist, H.H. Rugh and G. Vattay, *Scattering Theory - special issue*, CHAOS (1993).

[22.47] E.J. Heller, S. Tomsovic and A. Sepúlveda CHAOS **2**, *Periodic Orbit Theory - special issue*, 105, (1992).

[22.48] V.I. Arnold, *Geometrical Methods in the Theory of Ordinary Differential Equations*, (SpringerVerlag, New York 1983).

[22.49] 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).

[22.50] V.I. Arnold, *Geometrical Methods in the Theory of Ordinary Differential Equations* (Springer, New York 1983).

Exercises

22.1 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 $\Gamma(1+d/2) = \Gamma(d/2)d/2$.

22.2 Monodromy matrix from second variations of the action. Show that

$$D_{\perp j}/D'_{\perp j} = (\mathbf{1} - \mathbf{J}) \tag{22.14}$$

22.3 Jacobi gymnastics. Prove that the ratio of determinants in (K.28) can be expressed as

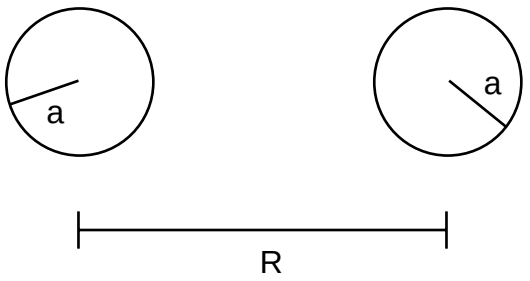
$$\frac{\det D'_{\perp j}(q_\parallel, 0, E)}{\det D_{\perp j}(q_\parallel, 0, E)} = \det \begin{pmatrix} I - \mathbf{J}_{qq} & -\mathbf{J}_{qp} \\ -\mathbf{J}_{pq} & I - \mathbf{J}_{pp} \end{pmatrix} = \det (1 - \mathbf{J}_j)\,, \tag{22.15}$$

where $\mathbf{J}_j$ is the monodromy matrix of the periodic orbit.

22.4 Quantum two-disk scattering. Compute the quasiclassical spectral determinant

$$Z(\varepsilon) = \prod_{p,j,l} \left(1 - \frac{t_p}{\Lambda_p^{j+2l}}\right)^{j+1}$$

for the two disk problem. Use the geometry



The full quantum mechanical version of this problem can be solved by finding the zeros in k for the determinant of the matrix

$$M_{m,n} = \delta_{m,n} + \frac{(-1)^n}{2} \frac{J_m(ka)}{H_n^{(1)}(ka)} \left(H_{m-n}^{(1)}(kR) + (-1)^n H_{m+n}^{(1)}(kR) \right) ,$$

where J_n is the nth 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; read chapter **??**; or read *Chaos* **2**, 79 (1992))

22.5 Pinball topological index. Upgrade your pinball simulator so that it compute the topological index for each orbit it finds.

22.6 Transport equations. Write the wave-function 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.

22.7 Zero length orbits. Derive the classical trace (7.1) 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.

Chapter 23

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 became 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

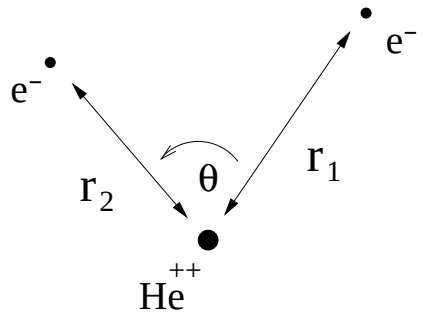


Figure 23.1: Coordinates for the helium three body problem in the plane.

from the insights we have today on the role chaos plays for helium and its quantum spectrum. We will introduce collinear helium in sect. 23.1 and discuss its dynamics in some detail. We will learn how to integrate and find periodic orbits and how to determine the relevant eigenvalues of the Jacobian matrix. We will explain in sect. 23.2 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. 23.2.1.

23.1 Classical dynamics of collinear helium

The full 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 $+2e$.

The helium electron-nucleus mass ratio $m_{he}/m_e = 1836$ is so large that we may work in the infinite nucleus mass approximation $m_{he} = \infty$, fixing the nucleus at the origin. Finite nucleus mass effects can be taken into account without any substantial difficulty. We are now left with two electrons moving in three spatial dimensions around the origin. The total angular momentum of the combined electron system is still conserved. In the special case of angular momentum $L = 0$, the electrons move in a fixed plane containing the nucleus. The three body problem can then be written in terms of three independent coordinates only, the electron-nucleus distances r_1 and r_2 and the inter-electron angle Θ, see fig. 23.1.

23.2 on p. 555

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 fig. 23.2, then there is no force acting on the electrons perpendicular to the

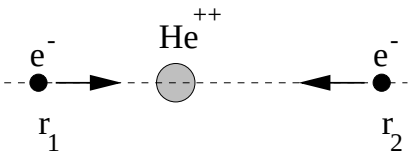


Figure 23.2: Collinear helium, with the two electrons on opposite sides of the nucleus.

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.

23.1.1 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

$$H = \frac{1}{2m_e}\left(p_1^2 + p_2^2\right) - \frac{2e}{r_1} - \frac{2e}{r_2} + \frac{e}{r_1 + r_2} = E\,, \tag{23.1}$$

where E is the total energy. 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

$$H = \frac{p_1^2}{2} + \frac{p_2^2}{2} - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_1 + r_2} = -1\,. \tag{23.2}$$

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. 23.2.

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 (21.32) which scales with energy as

$$S(E) = \oint \mathrm{d}\mathbf{q}(E) \cdot \mathbf{p}(E) = \frac{e^2 m_e^{1/2}}{(-E)^{1/2}}\, S, \tag{23.3}$$

with S being the action obtained from (23.2) for $E = -1$, and coordinates $\mathbf{q} = (r_1, r_2)$, $\mathbf{p} = (p_1, p_2)$. For the Hamiltonian (23.2), the period of a cycle and its action are related by (21.34), $T_p = \frac{1}{2} S_p$.

23.1.2 Regularization of two–body collisions

Next, we have a closer look at the singularities in the Hamiltonian (23.2). 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 destiantion) 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

$$H = \frac{1}{2}p^2 - \frac{2}{x} = E\,. \tag{23.4}$$

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 intial point, we can immediately integrate

$$\frac{1}{2}\dot{x}^2 - \frac{2}{x} = 0$$

by means of separation of variables

$$\sqrt{x}dx = \sqrt{2}dt\,, \qquad x = (9/2)^{\frac{1}{3}}\, t^{\frac{2}{3}}\,, \tag{23.5}$$

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 (23.4) we choose the time transformation $dt = xd\tau$ which lifts the $|x| \to 0$ singularity in (23.4) and leads to a new Hamiltonian

$$\mathcal{H} = \frac{1}{2}xp^2 - 2 - Ex = 0. \tag{23.6}$$

The solution (23.5) is now parametrized by the fictitous time $d\tau$ through a pair of equations

$$x = \frac{1}{2}\tau^2\,, \qquad t = \frac{1}{6}\tau^3\,.$$

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 (23.5) now suggests a canonical transformation of form

$$x = Q^2\,, \quad p = \frac{P}{2Q} \tag{23.7}$$

which maps the Kepler problem into that of a harmonic oscillator with Hamiltonian

$$H(Q,P) = \frac{1}{8}P^2 - EQ^2 = 2, \tag{23.8}$$

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 (23.7), by introducing a new vector Q with property $r = |Q|^2$. In this simple 1-dimensional example the KS transformation can be implemented by

$$r_1 = Q_1^2\,, \qquad r_2 = Q_2^2\,, \qquad p_1 = \frac{P_1}{2Q_1}\,, \qquad p_2 = \frac{P_2}{2Q_2} \tag{23.9}$$

23.1 on p. 555

and reparametrization of time by $d\tau = dt/r_1 r_2$. The singular behavior in the original momenta at r_1 or $r_2 = 0$ is again compensated by stretching the time scale at these points. The Hamiltonian structure of the equations of motions with respect to the new time τ is conserved, if we consider the Hamiltonian

$$H_{ko} = \frac{1}{8}(Q_2^2 P_1^2 + Q_1^2 P_2^2) - 2R_{12}^2 + Q_1^2 Q_2^2 \left(1 + \frac{1}{R_{12}^2}\right) = 0 \tag{23.10}$$

with $R_{12} = (Q_1^2 + Q_2^2)^{1/2}$. The equations of motion now have the form

$$\begin{aligned} \dot{P}_1 &= 2Q_1 \left[2 - \frac{P_2^2}{8} - Q_2^2\left(1 + \frac{Q_2^2}{R_{12}^4}\right)\right]; \qquad \dot{Q}_1 = \frac{1}{4}P_1 Q_2^2 \\ \dot{P}_2 &= 2Q_2 \left[2 - \frac{P_1^2}{8} - Q_1^2\left(1 + \frac{Q_1^2}{R_{12}^4}\right)\right]; \qquad \dot{Q}_2 = \frac{1}{4}P_2 Q_1^2. \end{aligned} \tag{23.11}$$

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 (23.11) are singular only at the triple collision $R_{12} = 0$, that is, when both electrons hit the nucleus at the same time.

The new coordinates and the Hamiltonian (23.10) 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.

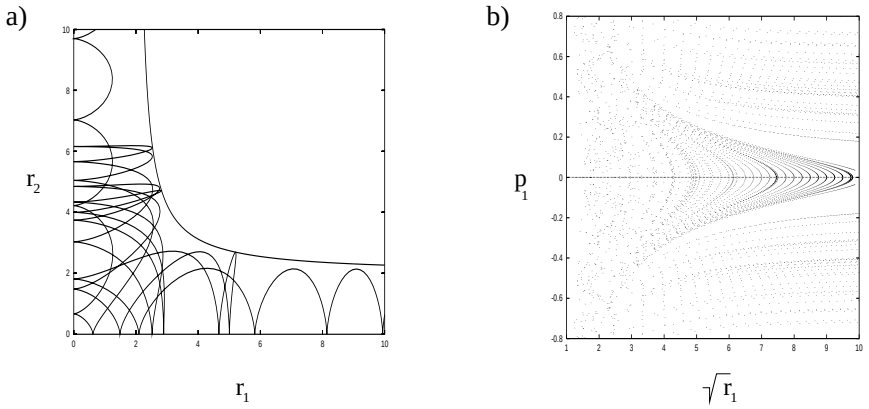


Figure 23.3: a) A typical trajectory in the r_1 – r_2 plane; the trajectory enters here along the r_1 axis and escapes to infinity along the r_2 axis; b) Poincaré map (r_2=0) for collinear helium. Strong chaos prevails for small r_1 near the nucleus.

23.1.3 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 fig. 23.3(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 fig. 23.3(b). We plot here the coordinate and momentum of the outer electron whenever the inner particle hits the nucleus, that is, 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

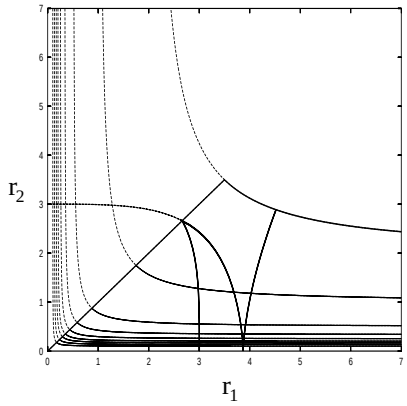


Figure 23.4: The cycle 011 in the fundamental domain $r_1 \geq r_2$ (full line) and in the full domain (dashed line).

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.

Reflection symmetry

The Hamiltonian (23.1) is invariant with respect to electron–electron exchange; this symmetry corresponds to the mirror symmetry of the potential along the line $r_1 = r_2$, fig. 23.4. 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 17, the dynamics in the fundamental domain is the key to the factorization of spectral determinants, to be implemented here in (23.22). Note also the similarity between the fundamental domain of the collinear potential fig. 23.4, and the fundamental domain fig. 10.4(b) in the 3–disk system, a simpler problem with the same binary symbolic dynamics.

in depth:
sect. 17.6, p. 397

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 (23.2) 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, that is, 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.

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. 11.5.2. 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 multi-dimensional 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 fig. 23.5. 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 (that is separable) limit of the dynamics and is thus marginally stable. The existence of this orbit is also related to intermittent behavior generating the quasi–regular dynamics for large r_1 that we have already noted in fig. 23.3(b).

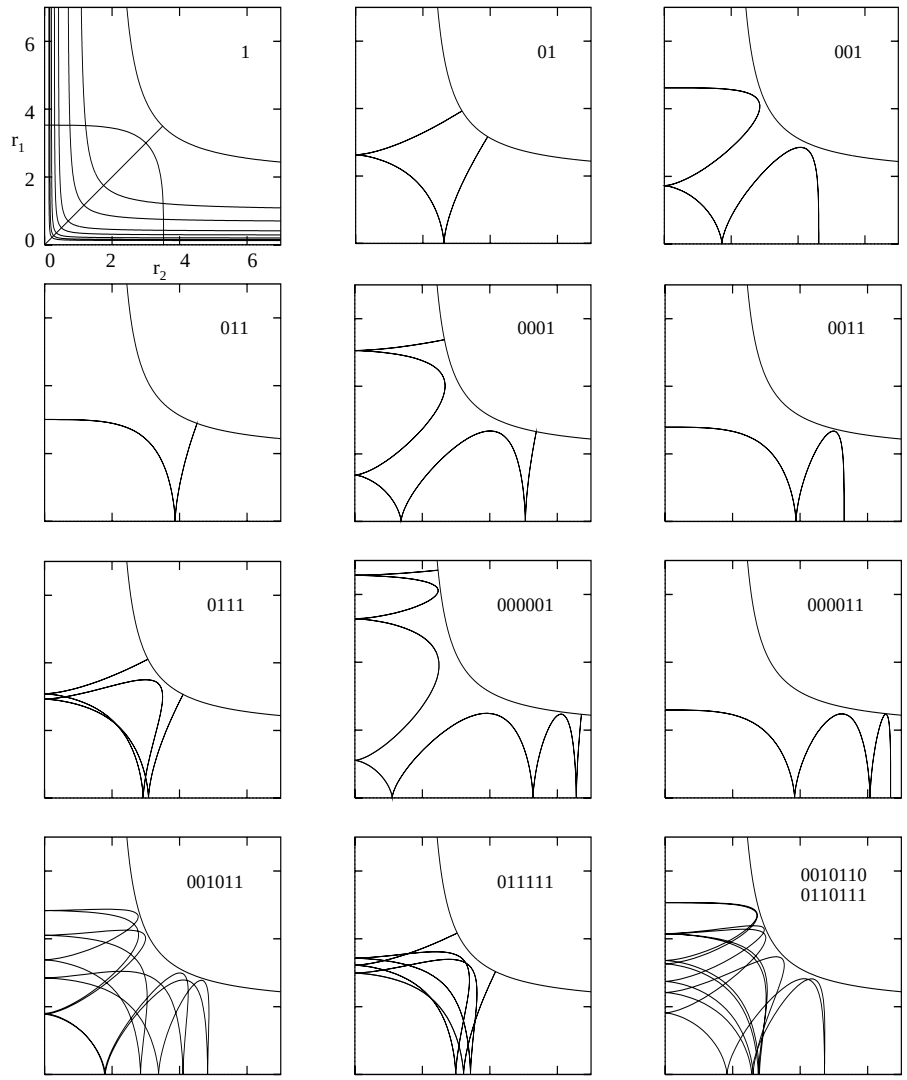


Figure 23.5: Some of the shortest cycles in collinear helium. The classical collinear electron motion is bounded by the potential barrier $-1 = -2/r_1 - 2/r_2 + 1/(r_1 + r_2)$ and the condition $r_i \geq 0$. The orbits are shown in the full r_1–r_2 domain, the itineraries refers to the dynamics in the $r_1 \geq r_2$ fundamental domain. The last figure, the 14-cycle 00101100110111, is an example of a typical cycle with no symmetry.

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 fig. 23.5 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 (that is 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 fig. 23.5;
- 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 fig. 23.5;
- 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 fig. 23.5.

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 fig. 23.5. Their determination would require a two-parameter search. The two cycles are mapped onto each other by time reversal symmetry, that is, 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 (23.11) 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 Jacobian matrix along a trajectory in order to calculate stability indices. We will provide the main equations in the next section, with the details of the derivation relegated to the appendix C.2.

23.5 on p. 556

23.1.4 Local coordinates, Jacobian matrix

In this section, we will derive the equations of motion for the Jacobian matrix along a collinear helium trajectory. The Jacobian matrix is 4-dimensional; the two trivial eigenvectors corresponding to the conservation of energy and displacements along a trajectory can, however, be projected out by suitable orthogonal coordinates transformations, see appendix C. We will give the transformation to local coordinates explicitly, here for the regularized coordinates (23.9), and state the resulting equations of motion for the reduced $[2 \times 2]$ Jacobian matrix.

The vector locally parallel to the trajectory is pointing in the direction of the phase space velocity (4.33)

$$v_m = \dot{x}_m(t) = \omega_{mn} \frac{\partial H}{\partial x_n} = (H_{P_1}, H_{P_2}, -H_{Q_1}, -H_{Q_2})^T,$$

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 (23.10)

$$\gamma = \nabla H = (H_{Q_1}, H_{Q_2}, H_{P_1}, H_{P_2})^T .$$

By symmetry $v_m \gamma_m = \omega_{mn} \frac{\partial H}{\partial x_n} \frac{\partial H}{\partial x_m}$, so the two vectors are orthogonal.

Next, we consider the orthogonal matrix

$$\begin{aligned} \mathbf{O}_{\gamma(t)} &= (\gamma_1, \gamma_2, \gamma/R, v) \qquad (23.12) \\ &= \begin{pmatrix} -H_{P_2}/R & H_{Q_2} & H_{Q_1}/R & H_{P_1} \\ H_{P_1}/R & -H_{Q_1} & H_{Q_2}/R & H_{P_2} \\ -H_{Q_2}/R & -H_{P_2} & H_{P_1}/R & -H_{Q_1} \\ H_{Q_1}/R & H_{P_1} & H_{P_2}/R & -H_{Q_2} \end{pmatrix} \end{aligned}$$

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 (γ, v). The vectors $\gamma_{1,2}$ are phase space vectors perpendicular to the trajectory and to the energy manifold in the 4-dimensional phase space of collinear helium. The Jacobian matrix (4.5) rotated to the local coordinate system by $\mathbf{O}$ then has the form

23.6 on p. 556

$$\mathbf{m} = \begin{pmatrix} m_{11} & m_{12} & * & 0 \\ m_{21} & m_{22} & * & 0 \\ 0 & 0 & 1 & 0 \\ * & * & * & 1 \end{pmatrix}, \quad \mathbf{J} = \mathbf{O^T m O} \sigma$$

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 Jacobian matrix $\mathbf{m}$ are given by

$$\dot{\mathbf{m}} = \mathbf{l}(t)\mathbf{m}(t), \tag{23.13}$$

with $\mathbf{m}(0) = \mathbf{1}$. The matrix $\mathbf{l}$ depends on the trajectory in phase space and has the form

$$\mathbf{l} = \begin{pmatrix} l_{11} & l_{12} & * & 0 \\ l_{21} & l_{22} & * & 0 \\ 0 & 0 & 0 & 0 \\ * & * & * & 0 \end{pmatrix},$$

where the relevant matrix elements l_{ij} are given by

$$\begin{aligned} l_{11} &= \frac{1}{R}[2H_{Q_1Q_2}(H_{Q_2}H_{P_1} + H_{Q_1}H_{P_2}) \\ &\quad +(H_{Q_1}H_{P_1} - H_{Q_2}H_{P_2})(H_{Q_1Q_1} - H_{Q_2Q_2} - H_{P_1P_1} + H_{P_2P_2})] \\ l_{12} &= -2H_{Q_1Q_2}(H_{Q_1}H_{Q_2} - H_{P_1}H_{P_2}) \\ &\quad +(H_{Q_1}^2 + H_{P_2}^2)(H_{Q_2Q_2} + H_{P_1P_1}) + (H_{Q_2}^2 + H_{P_1}^2)(H_{Q_1Q_1} + H_{P_2P_2}) \\ l_{21} &= \frac{1}{R^2}[2(H_{Q_1P_2} + H_{Q_2P_1})(H_{Q_2}H_{P_1} + H_{Q_1}H_{P_8}) \\ &\quad -(H_{P_1}^2 + H_{P_2}^2)(H_{Q_1Q_1} + H_{Q_2Q_2}) - (H_{Q_1}^2 + H_{Q_2}^2)(H_{P_1P_1} + H_{P_2P_2})] \\ l_{22} &= -l_{11}\,. \end{aligned} \tag{23.14}$$

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.

23.1.5 Getting ready

Now everything is in place: the regularized equations of motion can be implemented in a Runge–Kutta or any other integration scheme to calculate trajectories. We have a symbolic dynamics and know how many cycles there are and how to find them (at least up to symbol length 5). We know how to compute the Jacobian matrix whose eigenvalues enter the semiclassical spectral determinant (22.7). By (21.34) the action S_p is proportional to the period of the orbit, $S_p = 2T_p$.

p	$S_p/2\pi$	$\ln\|\Lambda_p\|$	σ_p	m_p
1	1.82900	0.6012	0.5393	2
01	3.61825	1.8622	1.0918	4
001	5.32615	3.4287	1.6402	6
011	5.39451	1.8603	1.6117	6
0001	6.96677	4.4378	2.1710	8
0011	7.04134	2.3417	2.1327	8
0111	7.25849	3.1124	2.1705	8
00001	8.56618	5.1100	2.6919	10
00011	8.64306	2.7207	2.6478	10
00101	8.93700	5.1562	2.7291	10
00111	8.94619	4.5932	2.7173	10
01011	9.02689	4.1765	2.7140	10
01111	9.07179	3.3424	2.6989	10
000001	10.13872	5.6047	3.2073	12
000011	10.21673	3.0323	3.1594	12
000101	10.57067	6.1393	3.2591	12
000111	10.57628	5.6766	3.2495	12
001011	10.70698	5.3251	3.2519	12
001101	10.70698	5.3251	3.2519	12
001111	10.74303	4.3317	3.2332	12
010111	10.87855	5.0002	3.2626	12
011111	10.91015	4.2408	3.2467	12

Table 23.1: Action S_p (in units of 2π), Lyapunov exponent $|\Lambda_p|/T_p$ for the motion in the collinear plane, winding number σ_p for the motion perpendicular to the collinear plane, and the topological index m_p for all fundamental domain cycles up to topological length 6.

There is, however, still a slight complication. Collinear helium is an invariant 4-dimensional subspace of the full helium phase space. If we restrict the dynamics to angular momentum equal zero, we are left with 6 phase space coordinates. That is not a problem when computing periodic orbits, they are oblivious to the other dimensions. However, the Jacobian matrix does pick up extra contributions. When we calculate the Jacobian matrix for the full problem, we must also allow for displacements out of the collinear plane, so the full Jacobian matrix for dynamics for $L = 0$ angular momentum is 6 dimensional. Fortunately, the linearized dynamics in and off the collinear helium subspace decouple, and the Jacobian matrix can be written in terms of two distinct $[2 \times 2]$ matrices, with trivial 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 Θ-coordinate in fig. 23.1. It turns out that the linearized dynamics in the Θ coordinate is stable, corresponding to a bending type motion of the two electrons. We will need the stability exponents for all degrees of freedom in evaluating the semiclassical spectral determinant in sect. 23.2.

The numerical values of the actions, stability exponents, stability angles, and topological indices for the shortest cycles are listed in table 23.1.4. These numbers, needed for the semiclassical quantization implemented in the next section, an also be helpful in checking your own calculations.

23.2 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 23.4.

23.2.1 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

$$E_N = -\frac{e^4 m_e}{\hbar^2}\frac{Z^2}{2N^2}\,, \tag{23.15}$$

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

$$E_{N,n} = -\frac{2}{N^2} - \frac{1}{2n^2} \qquad \text{with} \quad n > N. \tag{23.16}$$

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 (23.15) 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 (23.16) 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 fig. 23.6. 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

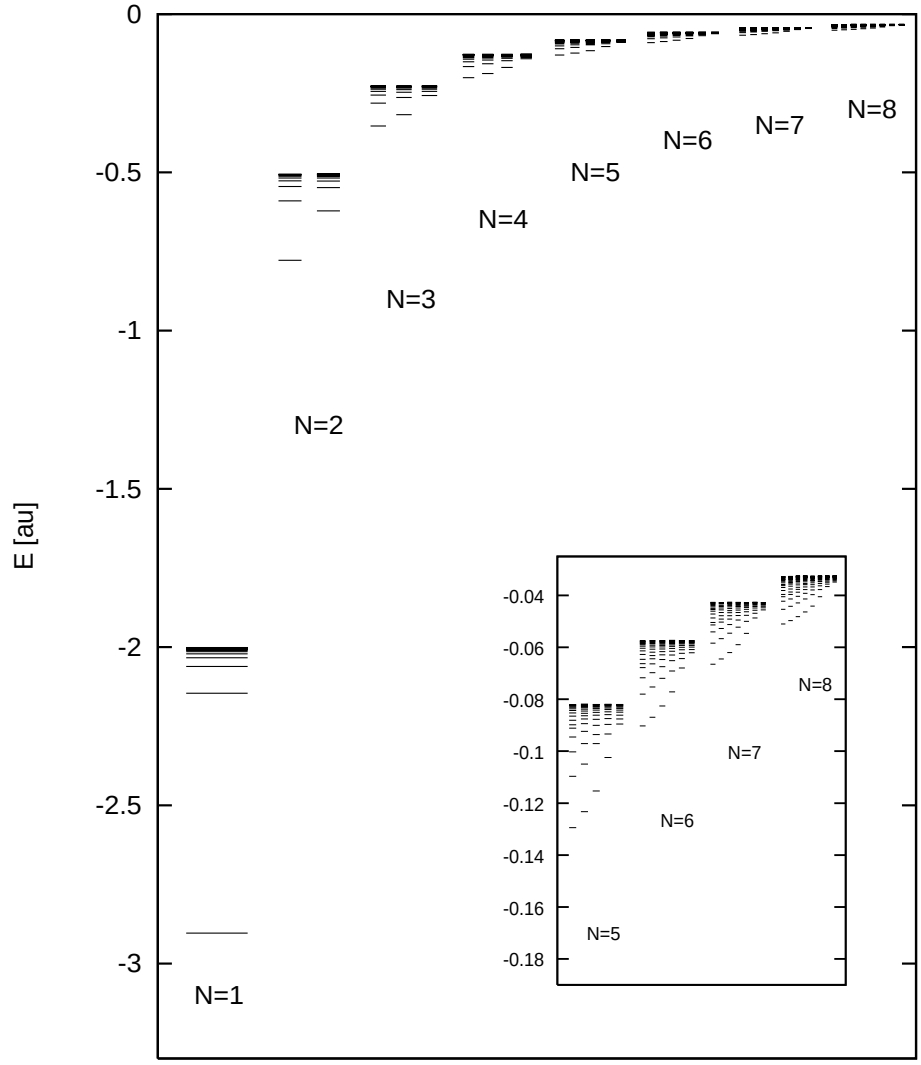


Figure 23.6: The exact quantum helium spectrum for $L = 0$. The energy levels denoted by bars have been obtained from full 3-dimensional quantum calculations [3].

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 fig. 23.6. We will see in the next section that this is indeed the case and that the collinear model holds down to the $N = 1$ bound state series, including even the ground state of helium! We will also find a semiclassical quantum number corresponding to the angular momentum l and show that the collinear model describes states for moderate angular momentum l as long as $l \ll N$. .

☞ remark 23.4

23.2.2 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 (22.7) 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. 23.1.1 which makes it possible to write the semiclassical spectral determinant in the form

$$\det\left(\hat{H}-E\right)_{sc} = \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{ir(sS_p - m_p\frac{\pi}{2})}}{\left(-\det\left(1-\mathbf{J}^r_{p\perp}\right)\right)^{1/2} \left|\det\left(1-\mathbf{J}^r_{p\parallel}\right)\right|^{1/2}}\right), \quad (23.17)$$

with the energy dependence absorbed into the variable

$$s = \frac{e^2}{\hbar}\sqrt{\frac{m_e}{-E}},$$

obtained by using the scaling relation (23.3) for the action. As explained in sect. 23.1.4, the fact that the $[4 \times 4]$ Jacobian matrix decouples into two $[2 \times 2]$ submatrices corresponding to the dynamics *in* the collinear space and *perpendicular* to it makes it possible to write the denominator in terms of a product of two determinants. Stable and unstable degrees of freedom enter the trace formula in different ways, reflected by the absence of the modulus sign and the minus sign in front of $\det\left(1-\mathbf{J}_\perp\right)$. 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 (22.7) is absent in (23.17). Collinear helium is an open system, that is 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 (22.13) we may proceed as in (8.9) by expanding the determinants in (23.17) in terms of the eigenvalues of the corresponding Jacobian matrices. The matrix representing displacements perpendicular to the collinear space has eigenvalues of the form $\exp(\pm 2\pi i\sigma)$, reflecting stable linearized dynamics. σ is the full winding number along the orbit in the stable degree of freedom, multiplicative under multiple repetitions of this orbit .The eigenvalues corresponding to the unstable dynamics along the collinear axis are paired as $\{\Lambda, 1/\Lambda\}$ with $|\Lambda| > 1$ and real. As in (8.9) and (22.13) we may thus write

$$\begin{aligned} &\left[-\det\left(1-\mathbf{J}^r_\perp\right)\left|\det\left(1-\mathbf{J}^r_\parallel\right)\right|\right]^{-1/2} \\ &= \left[-(1-\Lambda^r)(1-\Lambda^{-r})|(1-e^{2\pi i r\sigma})(1-e^{-2\pi i r\sigma})\right]^{-1/2} \\ &= \sum_{k,\ell=0}^{\infty} \frac{1}{|\Lambda^r|^{1/2}\Lambda^{rk}} e^{-ir(\ell+1/2)\sigma}\,. \end{aligned} \tag{23.18}$$

The $\pm$ sign corresponds to the hyperbolic/inverse hyperbolic periodic orbits with positive/negative eigenvalues Λ. Using the relation (23.19) we see that the sum over r in (23.17) is the expansion of the logarithm, so the semiclassical spectral determinant can be rewritten as a product over dynamical zeta functions, as in (8.9):

$$\det\left(\hat{H}-E\right)_{sc} = \prod_{k=0}^{\infty}\prod_{m=0}^{\infty}\zeta^{-1}_{k,m} = \prod_{k=0}^{\infty}\prod_{m=0}^{\infty}\prod_{p}(1-t_p^{(k,m)})\,, \tag{23.19}$$

where the cycle weights are given by

$$t_p^{(k,m)} = \frac{1}{|\Lambda|^{1/2}\Lambda^k} e^{i\left(sS_p - m_p\frac{\pi}{2} - 4\pi(\ell+1/2)\sigma_p\right)}\,, \tag{23.20}$$

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 σ. The values for the actions, winding numbers and stability indices of the shortest cycles in collinear helium are listed in table 23.1.4.

The integer indices ℓ and k play very different roles in the semiclassical spectral determinant (23.19). A linearized approximation of the flow along a cycle corresponds to a harmonic approximation of the potential in the vicinity of the trajectory. Stable motion corresponds to a harmonic oscillator potential, unstable motion to an inverted harmonic oscillator. The index ℓ which contributes

as a phase to the cycle weights in the dynamical zeta functions can therefore be interpreted as a harmonic oscillator quantum number; it corresponds to vibrational modes in the Θ coordinate and can in our simplified picture developed in sect. 23.2.1 be related to the quantum number $l = l_1 = l_2$ representing the single particle angular momenta. Every distinct ℓ value corresponds to a full spectrum which we obtain from the zeros of the semiclassical spectral determinant $1/\zeta_\ell$ keeping ℓ fixed. The harmonic oscillator approximation will eventually break down with increasing off-line excitations and thus increasing ℓ. The index k corresponds to 'excitations' along the unstable direction and can be identified with local resonances of the inverted harmonic oscillator centered on the given orbit. The cycle contributions $t_p^{(k,m)}$ decrease exponentially with increasing k. Higher k terms in an expansion of the determinant give corrections which become important only for large negative imaginary s values. As we are interested only in the leading zeros of (23.19), that is 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. 23.1.3. 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. 17.1.2 and 17.5, 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

$$1/\zeta^{(\ell)} = \prod_a (1 - t_a)^2 \prod_{\tilde{s}} (1 - t_{\tilde{s}}^2)\,. \tag{23.21}$$

Here, the first product is taken over all asymmetric prime cycles, that is 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 (23.21) are the weights of fundamental domain cycles. The C_2 symmetry now leads to the factorization of (23.21) $1/\zeta = \zeta_+^{-1}\zeta_-^{-1}$, with

$$\begin{aligned} 1/\zeta_+^{(\ell)} &= \prod_a (1 - t_a) \prod_{\tilde{s}} (1 - t_{\tilde{s}})\,, \\ 1/\zeta_-^{(\ell)} &= \prod_a (1 - t_a) \prod_{\tilde{s}} (1 + t_{\tilde{s}})\,, \end{aligned} \tag{23.22}$$

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. 23.1.3. 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 (23.22) with a negative or a positive sign, depending on the symmetry subspace under consideration.

23.2.3 Cycle expansion results

So far we have established a factorized form of the semiclassical spectral determinant and have thereby picked up two *good quantum numbers*; the quantum number m has been identified with an excitation of the bending vibrations, the exchange symmetry quantum number ± 1 corresponds to states being symmetric or antisymmetric with respect to the electron-electron exchange. We may now start writing down the binary cycle expansion (13.5) 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 16. 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 23.5

$$
\begin{aligned}
1/\zeta^{\ell}(s) = \quad 1 \quad & -t_1^{(\ell)} - t_{01}^{(\ell)} - [t_{001}^{(\ell)} + t_{011}^{(\ell)} - t_{01}^{(\ell)} t_1^{(\ell)}] \\
& -[t_{0001}^{(\ell)} + t_{0011}^{(\ell)} - t_{001}^{(\ell)} t_1^{(\ell)} + t_{0111}^{(\ell)} - t_{011}^{(\ell)} t_1^{(\ell)}] - \dots \ . \qquad (23.23)
\end{aligned}
$$

The weights $t_p^{(\ell)}$ are given in (23.19), 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 23.1.4. To get reacquainted with the cycle expansion formula (23.23), 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

$$E_{m,N} = -\frac{(S_1/2\pi)^2}{[m + \frac{1}{2} + 2(N + \frac{1}{2})\sigma_1]^2}, \qquad m, N = 0, 1, 2, \ldots \,, \tag{23.24}$$

with $S_1/2\pi = 1.8290$ for the action and $\sigma_1 = 0.5393$ for the winding number, see table 23.1.4, the 1 cycle in the fundamental domain. This cycle can be described as the *asymmetric stretch* orbit, see fig. 23.5. The additional quantum number N in (23.24) corresponds to the principal quantum number defined in sect. 23.2.1. The states described by the quantization condition (23.24) 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 fig. 23.6. The simple formula (23.24) gives already a rather good estimate for the ground state of helium! Results obtained from (23.24) are tabulated in table 23.2, 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 (23.23), 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 23.2 for symmetric states obtained from truncations of the cycle expansion of $1/\zeta_+$.

23.7 on p. 556

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 (23.22) reveals states in the Rydberg series which are to the right of the energetically lowest series in fig. 23.6.

Commentary

Remark **23.1** 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

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 23.2: Collinear helium, real part of the symmetric subspace resonances obtained by a cycle expansion (23.23) 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.

a plane, is due to Kustaanheimo [12] who realized that the correct higher-dimensional generalization of the "square root removal" trick (23.7), 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 **23.2** 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 labelled 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, that is by those trajectories which start in or end at the singular point $r_1 = r_2 = 0$. See also ref. [2].

Remark **23.3** 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 **23.4** Helium quantum numbers. The classification of the helium states in terms of single electron quantum numbers, sketched in sect. 23.2.1, prevailed until the 1960's; a growing discrepancy between experimental results and theoretical predictions made it necessary to refine this picture. In particular, the different Rydberg series sharing a given N-quantum number correspond, roughly speaking, to a quantization of the inter electronic angle Θ, see fig. 23.1, 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. 23.2.1 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 fig. 23.6, and contain the energetically highest states for given N, n quantum numbers, see also remark 23.5. A detailed account of the historical development as well as a modern interpretation of the spectrum can be found in ref. [1].

Remark **23.5** 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. 23.2 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 fig. 23.6. Details can be found in refs. [10, 11].

In order to obtain the Rydberg series structure of the spectrum, that is 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 (23.19), and cycle expansion of indeed yield all Rydbergstates up the various ionization thresholds, see ref. [4].

For a comprehensive overview on spectra of two-electron atoms and semiclassical treatments ref. [1].

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 ℓ. We thus understood the origin of the different Rydberg series depicted in fig. 23.6, 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.

References

[23.1] G. Tanner, J-M. Rost and K. Richter, Rev. Mod. Phys. **72**, 497 (2000).

[23.2] K. Richter, G. Tanner, and D. Wintgen, Phys. Rev. A **48**, 4182 (1993).

[23.3] Bürgers A., Wintgen D. and Rost J. M., J. Phys. B **28**, 3163 (1995).

[23.4] G. Tanner and D. Wintgen Phys. Rev. Lett. **75** 2928 (1995).

[23.5] P. Kustaanheimo and E. Stiefel, *J. Reine Angew. Math.* **218**, 204 (1965).

[23.6] E.L. Steifel and G. Scheifele, *Linear and regular celestial mechanics* (Springer, New York 1971).

[23.7] G.S. Ezra, K. Richter, G. Tanner and D. Wintgen, J. Phys. B **24**, L413 (1991).

[23.8] D. Wintgen, K. Richter and G. Tanner, CHAOS **2**, 19 (1992).

[23.9] R. Blümel and W. P. Reinhardt, *Directions in Chaos Vol 4*, eds. D. H. Feng and J.-M. Yuan (World Scientific, Hongkong), 245 (1992).

[23.10] K. Richter and D. Wintgen, J. Phys. B **24**, L565 (1991).

[23.11] D. Wintgen and K. Richter, Comments At. Mol. Phys. **29**, 261 (1994).

[23.12] P. Kustaanheimo, *Ann. Univ. Turku, Ser. AI.,* **73** (1964).

[23.13] T. Levi-Civita, *Opere mathematische* **2** (1956).

[23.14] H. Hopf, *Math. Ann.* **104** (1931).

Exercises

23.1 Kustaanheimo–Stiefel transformation. Check the Kustaanheimo–Stiefel regularization for collinear helium; derive the Hamiltonian (23.10) and the collinear helium equations of motion (23.11).

23.2 Helium in the plane. Starting with the helium Hamiltonian in the infinite nucleus mass approximation $m_{he} = \infty$, and angular momentum $L = 0$, show that the three body problem can be written in terms of three independent coordinates only, the electron-nucleus distances r_1 and r_2 and the inter-electron angle Θ, see fig. 23.1.

23.3 Helium trajectories. Do some trial integrations of the collinear helium equations of motion (23.11). 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 23.4. Most trajectories will run away, do not be surprised - the classical collinear helium is unbound. Try to guess approximately the shortest cycle of fig. 23.4.

23.4 A Poincaré section for collinear Helium. Construct a Poincaré section of fig. 23.3b that reduces the helium flow to a map. Try to delineate regions which correspond to finite symbol sequences, that is initial conditions that follow the same topological itinerary in the fig. 23.3a 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 23.5.

23.5 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 (23.23).

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.

23.6 Collinear helium cycle stabilities. Compute the eigenvalues for the cycles you found in exercise 23.5, as described in sect. 23.1.4. You may either integrate the reduced 2×2 matrix using equations (23.13) together with the generating function $\mathbf{l}$ given in local coordinates by (23.14) or integrate the full 4×4 Jacobian matrix, see sect. 15.1. 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 23.1.4; you should get the actions and Lyapunov exponents right, but topological indices and stability angles we take on faith.

23.7 Helium eigenenergies. Compute the lowest eigenenergies of singlet and triplet states of helium by substituting cycle data into the cycle expansion (23.23) for the symmetric and antisymmetric zeta functions (23.22). 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 fig. 13.1, can yield informed guesses. Better way would be to find the zeros by Newton method, sect. 13.1.4. How close are you to the cycle expansion and quantum results listed in table 23.2? You can find more quantum data in ref. [3].

Chapter 24

Diffraction distraction

(N. Whelan)

Diffraction effects characteristic to scattering off wedges are incorporated into the periodic orbit theory.

24.1 Quantum eavesdropping

As noted in chapter 23, 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 chapter **??**. 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

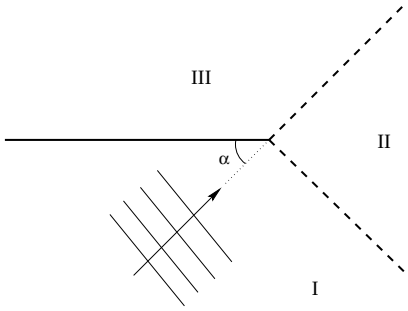


Figure 24.1: Scattering of a plane wave off a half line.

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.

We start by solving the problem of the scattering of a plane wave off a half line (see fig. 24.1). This is the local problem whose solution we will use to construct a global solution of more complicated geometries. We define the vertex to be the origin and launch a plane wave at it from an angle α. What is the total field? This is a problem solved by Sommerfeld 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 semi-classical 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

$$v(r,\psi) = Ae^{-ikr\cos\psi} \tag{24.1}$$

where $\psi = \phi - \alpha$ and ϕ is the angular coordinate. The total field is then given by the method of images as

$$v_{\rm tot} = v(r,\phi-\alpha) - v(r,\phi+\alpha), \tag{24.2}$$

where the negative sign ensures that the boundary condition of zero field on the line is satisfied.

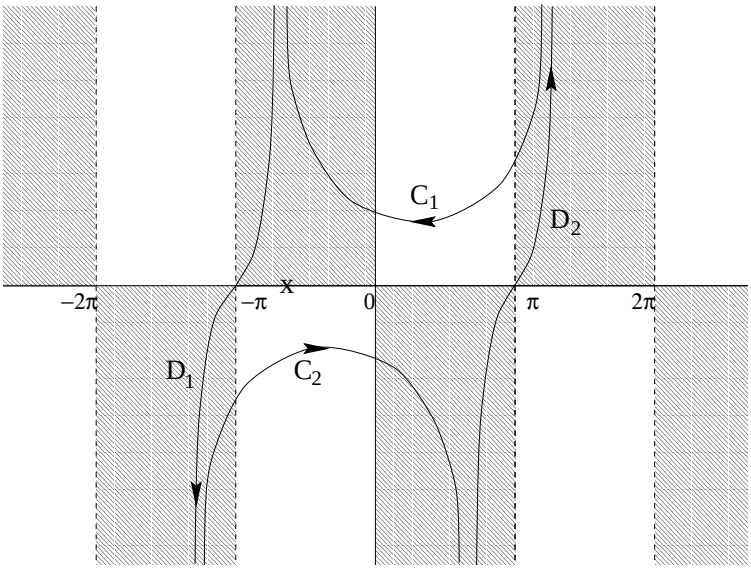


Figure 24.2: The contour in the complex β plane. The pole is at $\beta = -\psi$ (marked by $\times$ in the figure) and the integrand approaches zero in the shaded regions as the magnitude of the imaginary part of β approaches infinity.

Sommerfeld then argued that $v(r, \psi)$ can also be given a complex integral representation

$$v(r, \psi) = A \int_C d\beta f(\beta, \psi) e^{-ikr \cos \beta}. \tag{24.3}$$

This is certainly correct if the function $f(\beta, \psi)$ has a pole of residue $1/2\pi i$ at $\beta = -\psi$ and if the contour C encloses that pole. One choice is

$$f(\beta, \psi) = \frac{1}{2\pi} \frac{e^{i\beta}}{e^{i\beta} - e^{-i\psi}}. \tag{24.4}$$

(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 fig. 24.2. 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 2π. However, they are traversed in an opposite sense and cancel, so our contour consists of just the sections C_1 and C_2. The motivation for expressing the solution in this complicated manner should become clear soon.

What have we done? We extended the space under consideration by a factor of two and then constructed a solution by assuming that there is also a source

in the unphysical space. We superimpose the solutions from the two sources and at the end only consider the solution in the physical space to be meaningful. Furthermore, we expressed the solution as a contour integral which reflects the 2π periodicity of the problem. The half line scattering problem follows by analogy.

Whereas for the full line the field is periodic in 2π, for the half line it is periodic in 4π. This can be seen by the fact that the field can be expanded in a series of the form $\{\sin(\phi/2), \sin(\phi), \sin(3\phi/2), \cdots\}$. As above, we extend the space by thinking of it as two sheeted. The physical sheet is as shown in fig. 24.1 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 as

$$v_{\text{tot}} = u(r, \phi - \alpha) - u(r, \phi + \alpha). \tag{24.5}$$

where $u(r, \psi)$ is a 4π-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 (24.3) with the same contour $C_1 + C_2$ but with the 4π periodicity accounted for by replacing equation (24.4) with

$$f(\beta, \psi) = \frac{1}{4\pi} \frac{e^{i\beta/2}}{e^{i\beta/2} - e^{-i\psi/2}}. \tag{24.6}$$

(We divide by 4π rather than 2π so that the residue is properly normalized.) The integral (24.3) 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 (24.4). Therefore, this ansatz is a solution of the equation and satisfies all boundary conditions and therefore constitutes a valid solution. By uniqueness this is the only solution.

In order to further understand this solution, it is useful to massage the contour. Depending on ϕ there may or may not be a pole between $\beta = -\pi$ and $\beta = \pi$. In region I, both functions $u(r, \phi \pm \alpha)$ have poles which correspond to the

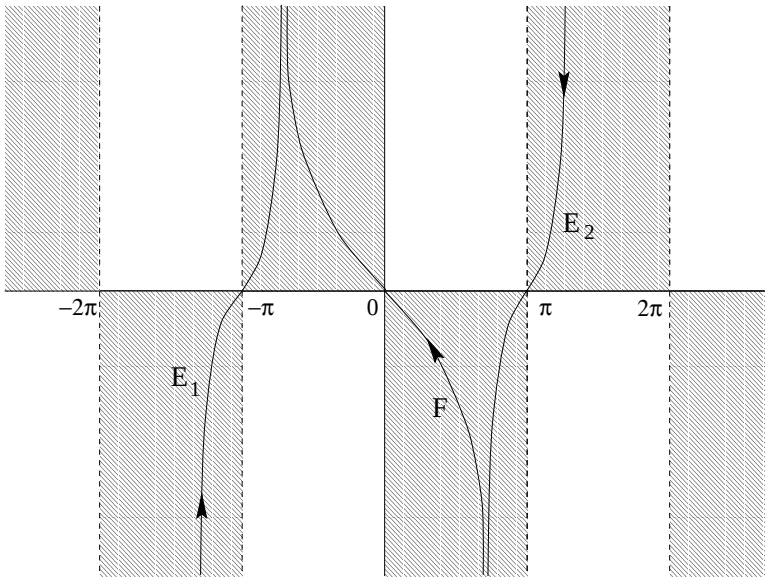


Figure 24.3: The contour used to evaluate the diffractive field after the contribution of possible poles has been explicitly evaluated. The curve F is traversed twice in opposite directions and has no net contribution.

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 fig. 24.2.

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 fig. 24.2 but with a reversed sense. Since the integrand is no longer 2π periodic, the contributions from these curves no longer cancel. We evaluate both stationary phase integrals to obtain

$$u(r, \psi) \approx -A \frac{e^{i\pi/4}}{\sqrt{8\pi}} \sec(\psi/2) \frac{e^{ikr}}{\sqrt{kr}} \tag{24.7}$$

so that the total diffracted field is

$$v_{\text{diff}} = -A \frac{e^{i\pi/4}}{\sqrt{8\pi}} \left(\sec\left(\frac{\phi - \alpha}{2}\right) - \sec\left(\frac{\phi + \alpha}{2}\right) \right) \frac{e^{ikr}}{\sqrt{kr}}. \tag{24.8}$$

Note that this expression breaks down when $\phi \pm \alpha = \pi$. These angles correspond to the borders among the three regions of fig. 24.1 and must be handled more carefully - we can not do a stationary phase integral in the vicinity of a pole. However, the integral representation (24.3) and (24.6) is uniformly valid.

24.1 on p. 573

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

$$G_{\rm f}(x_2, x_1, k) = -\frac{i}{4} H_0^{(+)}(kd) \approx -\frac{1}{\sqrt{8\pi kd}} \exp\{i(kd + \pi/4)\}, \tag{24.9}$$

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 (24.8), 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

$$G_{\rm diff}(x, x', k) = G_{\rm f}(x, x_V, k) d(\theta, \theta') G_{\rm f}(x_V, x', k), \tag{24.10}$$

where, by comparison with equations (24.8) and (24.9), we have

$$d(\theta, \theta') = \sec\left(\frac{\theta - \theta'}{2}\right) - \sec\left(\frac{\theta + \theta'}{2}\right). \tag{24.11}$$

Here θ' is the angle to the source as measured from the vertex and θ is the angle to the receiver. They were denoted as α and ϕ previously. Note that there is a symmetry between the source and receiver as we expect for a time-reversal invariant process. Also the diffraction coefficient d does not depend on which face of the half line we use to measure the angles. As we will see, a very important property of $G_{\rm diff}$ is that it is a simple multiplicative combination of other semiclassical Green's functions.

24.2 on p. 573

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 (24.9) and (24.10) 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

$$G(x,x',k) = g(r,r',k,\theta'-\theta) - g(r,r',k,\theta'+\theta) \tag{24.12}$$

where (r,θ) and (r',θ') are the polar coordinates of the points x and x' as measured from the vertex and the angles are measured from either face of the wedge. The function g is given by

$$g(r,r',k,\psi) = \frac{i}{8\pi\nu}\int_{C_1+C_2} d\beta \frac{H_0^+(k\sqrt{r^2+r'^2-2rr'\cos\beta})}{1-\exp\left(i\frac{\beta+\psi}{\nu}\right)} \tag{24.13}$$

where $\nu = \gamma/\pi$ and γ is the opening angle of the wedge. (ie $\gamma = 2\pi$ in the case of the half plane). The contour $C_1 + C_2$ is the same as shown in fig. 24.2.

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

$$d(\theta,\theta') = -\frac{4\sin\frac{\pi}{\nu}}{\nu}\frac{\sin\frac{\theta}{\nu}\sin\frac{\theta'}{\nu}}{\left(\cos\frac{\pi}{\nu}-\cos\frac{\theta+\theta'}{\nu}\right)\left(\cos\frac{\pi}{\nu}-\cos\frac{\theta-\theta'}{\nu}\right)}, \tag{24.14}$$

which reduces to (24.11) when $\nu = 2$. Note that the diffraction coefficient vanishes identically if $\nu = 1/n$ where n is any integer. This corresponds to wedge angles of $\gamma = \pi/n$ (eg. n=1 corresponds to a full line and n=2 corresponds to a right angle). This demonstration is limited by the fact that it came from a leading order asymptotic expansion but the result is quite general. For such wedge angles, we can use the method of images (we will require $2n-1$ images in addition to the actual source point) to obtain the Green's function and there is no diffractive contribution to any order. 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.

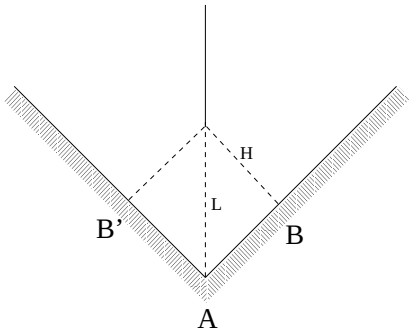


Figure 24.4: The billiard considered here. The dynamics consists of free motion followed by specular reflections off the faces. The top vertex induces diffraction while the bottom one is a right angle and induces two specular geometric reflections.

The integral representation is also useful because it allows us to consider geometries such that the angles are near the optical boundaries or the wedge angle is close to π/n. For these geometries the saddle point analysis leading to (24.14) 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.

24.2 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 fig. 24.4 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 appendix **??** for two dimensional scatterers. The poles have complex wave number 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 γ shown in fig. 24.5 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 (24.9) and (24.10). To leading order in y the lengths of the two geometric paths connecting P and V are $d_\pm = (L \pm x) + y^2/(L \pm x)^2/2$ so that the phase factor $ik(d_+ + d_-)$ equals $2ikL + iky^2/(L^2 - x^2)$. The trace integral involves integrating over all points P

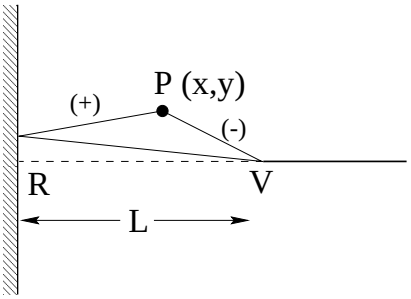


Figure 24.5: The dashed line shows a simple periodic diffractive orbit γ. Between the vertex V and a point P close to the orbit there are two geometric legs labeled $\pm$. The origin of the coordinate system is chosen to be at R.

and is

$$g_\gamma(k) \approx -2d_\gamma \frac{e^{i(2kL+\pi/2)}}{8\pi k} \int_0^L \frac{dx}{\sqrt{L^2-x^2}} \int_{-\infty}^{\infty} dy e^{\left(iky^2 \frac{L}{L^2-x^2}\right)}. \tag{24.15}$$

We introduced an overall negative sign to account for the reflection at the hard wall and multiplied by 2 to account for the two traversal senses, $VRPV$ and $VPRV$. In the spirit of stationary phase integrals, we have neglected the y dependence everywhere except in the exponential. The diffraction constant d_γ is the one corresponding to the diffractive periodic orbit. To evaluate the y integral, we use the identity

$$\int_{-\infty}^{\infty} d\xi e^{ia\xi^2} = e^{i\pi/4}\sqrt{\frac{\pi}{a}}, \tag{24.16}$$

and thus obtain a factor which precisely cancels the x dependence in the x integral. This leads to the rather simple result

$$g_\gamma \approx -\frac{il_\gamma}{2k}\left\{\frac{d_\gamma}{\sqrt{8\pi k l_\gamma}}\right\} e^{i(kl_\gamma+\pi/4)} \tag{24.17}$$

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 (24.13). It is valid in the vicinity of an optical boundary and also for wedges with opening angles close to π/n.

Consider a periodic diffractive orbit with n_γ reflections off straight hard walls and μ_γ diffractions each with a diffraction constant $d_{\gamma,j}$. The total length of the orbit $L_\gamma = \sum l_{\gamma,j}$ is the sum of the various diffractive legs and l_γ is the length of the corresponding primitive orbit. For such an orbit, (24.17) generalizes to

$$g_\gamma(k) = -\frac{il_\gamma}{2k}\left\{\prod_{j=1}^{\mu_\gamma} \frac{d_{\gamma,j}}{\sqrt{8\pi k l_{\gamma,j}}}\right\} \exp\{i(kL_\gamma + n_\gamma\pi - 3\mu_\gamma\pi/4)\}. \tag{24.18}$$

24.3 on p. 573

Each diffraction introduces a factor of $1/\sqrt{k}$ and multi-diffractive orbits are thereby suppressed.

If the orbit γ is primitive then $L_\gamma = l_\gamma$. If γ is the r'th repeat of a primitive orbit β we have $L_\gamma = rl_\beta$, $n_\gamma = rp_\beta$ and $\mu_\gamma = r\sigma_\beta$, where l_β, p_β and σ_β all refer to the primitive orbit. We can then write

$$g_\gamma = g_{\beta,r} = -\frac{il_\beta}{2k} t_\beta^r \tag{24.19}$$

where

$$t_\beta = \left\{ \prod_{j=1}^{\sigma_\beta} \frac{d_{\beta,j}}{\sqrt{8\pi k l_{\beta,j}}} \right\} \exp\left\{ i(kl_\beta + p_\beta\pi - 3\sigma_\beta\pi/4) \right\}. \tag{24.20}$$

It then makes sense to organize the sum over diffractive orbits as a sum over the primitive diffractive orbits and a sum over the repetitions

$$g_{\rm diff}(k) = \sum_\beta \sum_{r=1}^{\infty} g_{\beta,r} = -\frac{i}{2k} \sum_\beta l_\beta \frac{t_\beta}{1-t_\beta}. \tag{24.21}$$

We cast this as a logarithmic derivative (8.7) 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

$$g_{\rm diff}(k) \approx \frac{1}{2k}\frac{d}{dk} \left\{ \ln \prod_\beta (1-t_\beta) \right\}. \tag{24.22}$$

In the case that there are only diffractive periodic orbits - as in the geometry of fig. 24.4 - the poles of $g(k)$ are the zeros of a dynamical zeta function

$$1/\zeta(k) = \prod_\beta (1-t_\beta). \tag{24.23}$$

For geometric orbits, this function would be evaluated with a cycle expansion as discussed in chapter 13. However, here we can use the multiplicative nature of the weights t_β to find a closed form representation of the function using a Markov

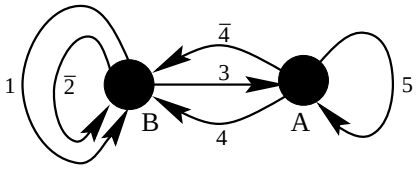


Figure 24.6: The two-node Markov graph with all the diffractive processes connecting the nodes.

graph, as in sect. 10.8.1. This multiplicative property of the weights follows from the fact that the diffractive Green's function (24.10) 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. 17.5 and 17.1.2) and we determine $1/\zeta_+$ and $1/\zeta_-$ separately using the ideas of symmetry decomposition of chapter 17.

In the Markov graph shown in fig. 24.6, we enumerate all processes. We start by identifying the fundamental domain as just the right half of fig. 24.4. 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. Thirdly, 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 fig. 24.6. 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. 17.5 we find

$$\begin{aligned} 1/\zeta_+ &= 1 - t_1 - t_{\bar{2}} - t_5 - t_3t_4 - t_3t_{\bar{4}} + t_5t_1 + t_5t_{\bar{2}}\,, \\ &= 1 - (t_1 + t_{\bar{2}} + t_5) - 2t_3t_4 + t_5(t_1 + t_{\bar{2}}) \end{aligned} \tag{24.24}$$

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. 17.3.1) and only affects the even resonances - the terms involving t_5 are absent from $1/\zeta_-$. The result is

$$1/\zeta_- = 1 - t_1 + t_{\bar{2}} - t_3t_4 + t_3t_{\bar{4}}\,,$$

$$= \quad 1-(t_1-t_{\bar{2}}). \tag{24.25}$$

24.4 on p. 573

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.

It now just remains to fix the weights. We use equation (24.20) but note that each weight involves just one diffraction constant. It is then convenient to define the quantities

$$u_A^2 = \frac{\exp\{i(2kL+2\pi)\}}{\sqrt{16\pi kL}} \qquad u_B^2 = \frac{\exp\{i(2kH+\pi)\}}{\sqrt{16\pi kH}}. \tag{24.26}$$

The lengths L and $H = L/\sqrt{2}$ are defined in fig. 24.4; we set $L = 1$ throughout. Bouncing inside the right angle at A corresponds to two specular reflections so that $p = 2$. We therefore explicitly include the factor $\exp(i2\pi)$ in (24.26) 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 (24.11) 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:

$$\begin{aligned} t_1 = sd_{BB}u_B^2 \quad t_{\bar{2}} &= sd_{B'B}u_B^2 \quad t_3 = t_4 = t_{\bar{4}} = sd_{AB}u_Au_B \\ t_5 &= sd_{AA}u_A^2. \end{aligned} \tag{24.27}$$

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 (24.25) this means that there are no odd resonances because 1 can never equal 0. For the even resonances equation (24.24) is an implicit equation for k which has zeros shown in fig. 24.7.

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

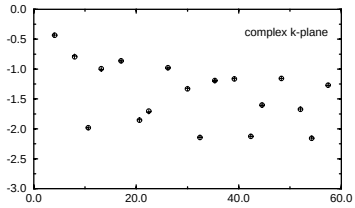


Figure 24.7: The even resonances of the wedge scatterer of fig. 24.4 plotted in the complex $k-$plane, with $L = 1$. The exact resonances are represented as circles and their semiclassical approximations as crosses.

24.5 on p. 574

to find so the dynamical zeta function approximation is particularly useful in that case.

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 fig. 24.4 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.

Commentary

Remark **24.1** 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 **24.2** 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 (24.8). 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 **24.3** Generalizations Consider the effect of replacing our half line by a wedge of angle γ_1 and the right angle by an arbitrary angle γ_2. If $\gamma_2 > \gamma_1$ and $\gamma_2 \geq \pi/2$ this is an open problem whose solution is given by equations (24.24) and (24.25) (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 (24.7) 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 **24.4** Diffractive Green's functions. The result (24.17) is proportional to the length of the orbit times the semiclassical Green's function (24.9) to go from the vertex back to itself along the classical path. The multi-diffractive formula (24.18) 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 **24.5** 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 23. 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].

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.

References

[24.1] A. Wirzba, CHAOS **2**, 77 (1992);

[24.2] 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).

[24.3] H. Primack, H. Schanz, U. Smilansky and I. Ussishkin, Phys. Rev. Lett. **76**, 1615 (1996).

[24.4] N. D. Whelan, Phys. Rev. E **51**, 3778 (1995).

[24.5] N. Pavloff and C. Schmit, Phys. Rev. Lett. **75**, 61 (1995).

[24.6] M. Sieber, N. Pavloff, C. Schmit, Phys. Rev. E **55**, 2279 (1997).

[24.7] H. Primack et. al., J. Phys. A **30**, 6693 (1997).

[24.8] N. D. Whelan, Phys. Rev. Lett. **76**, 2605 (1996).

[24.9] H. Bruus and N. D. Whelan, Nonlinearity, **9**, 1 (1996).

[24.10] P. Seba, Phys. Rev. Lett. **64**, 1855 (1990).

[24.11] P. E. Rosenqvist, N. D. Whelan and A. Wirzba, J. Phys. A **29**, 5441 (1996).

[24.12] M. Brack et. al., Chaos **5**, 317 (1995).

[24.13] S. M. Reimann et. al., Phys. Rev. A **53**, 39 (1996).

[24.14] A. Sommerfeld, Mathem. Ann. **47**, 317 (1896); *Optics* (Academic Press, New York 1954).

[24.15] H. S. Carslaw, Proc. London Math. Soc. (Ser. 1) **30**, 121 (1989); H. S. Carslaw, Proc. London Math. Soc. (Ser. 2) **18**, 291 (1920).

[24.16] J. B. Keller, J. Appl. Phys. **28**, 426 (1957).

[24.17] A. Voros, J. Phys. A **21**. 685 (1988).

[24.18] see for example, D. Ruelle, *Statistical Mechanics, Thermodynamic Formalism* (Addison-Wesley, Reading MA, 1978).

[24.19] see for example, P. Grassberger, Z. Naturforsch. **43a**, 671 (1988).

[24.20] P. Cvitanović and B. Eckhardt, Nonlinearity **6**, 277 (1993).

[24.21] P. Exner, P. Seba and P. Stovicek, Czech J. Phys **B39**, 1181 (1989).

[24.22] Hua Wu and D. W. L. Sprung, J. Appl. Phys. **72**, 151 (1992).

[24.23] 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).

[24.24] 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).

Exercises

(N. Whelan)

24.1 Stationary phase integral. Evaluate the two stationary phase integrals corresponding to contours E_1 and E_2 of fig. 24.3 and thereby verify (24.7).

24.2 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

$$d = \frac{2\pi}{\log\left(\frac{2}{ka}\right) - \gamma_e + i\frac{\pi}{2}} \tag{24.28}$$

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.

24.3 Several diffractive legs. Derive equation (24.18). 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.

24.4 Unsymmetrized dynamical zeta function. Assume you know nothing about symmetry decomposition. Construct the three node Markov diagram for fig. 24.1 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?

24.5 Three point scatterers. Consider the limiting case of the three disk game of pinball of fig. 1.1 where the disks are very much smaller than their spacing R. Use the results of exercise 24.2 to construct the desymmetrized dynamical zeta functions, as in sect. 17.6. 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].)

Epilogue

> Nowdays, 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)

24.3 Cycles as the skeleton of chaos

> 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 wind down this chatty chapter 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 superexponentially 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 dynam-

ics 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 Jacobian 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", that is 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 (see sect. **??**), for which there is no guarantee that they converge to the correct asymptotic value in any finite time numerical computation.

?? on p. ??

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 chapter I.1.2, and for smooth flows the convergence of cycle expansions can even be super-exponential.

The cycle expansions such as (13.5) outperform the pedestrian methods such as extrapolations from the finite cover sums (14.2) 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 nth level sum (14.2) 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 (14.2) the limit has to be estimated by numerical extrapolations. And, crucially, the higher terms in the cycle expansion

(13.5) 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. [9].

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 assignement 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

t_p
    cycle weigth, 150

, 79, 203
Schrödinger, E, 105

abscissa
    absolute conv., 309
    conditional conv., 309
accelerator mode, 411
action, 489, 500
    helium, 532
    relation to period, 541
admissible
    periodic points, 243
    trajectories, number of, 239
Airy function, 510
    at a bifurcation, 510
alphabet, 198
alternating binary tree, 235
analyticity
    domain, 147
Anosov flows, 221
antiharmonic extension, 630
arc, 203
area
    preserving map, 64
Artin-Mazur zeta function, 247
attractor, 36
    basin, 36
    Hénon, 70
    strange, 36, 40, 126
autonomous flow, 39, 275
average
    space, 103
    time, 103
averages
    chaotic, 363
averaging, 29
    space, 120
    time, 118
Axiom A
    systems, 182

baker's map, 66
Balmer spectrum, 480
basin of attraction, 36
Bernoulli
    polynomials, 176
    shift, 175
bi-infinite itinerary, 200
bifurcation
    Airy function approximation, 510
    bizarre, 629
    generic, 67
    Hopf, 425
    saddle-node, 66
billiard
    map, 65
billiards, 64, 66
    stability, 83
binary
    symbolic dynamics
        collinear helium, 537
    tree, alternating, 235
Birkhoff
    coordinates, 11
    ergodic theorem, 104
block
    finite sequence, 201
    pruning, 202
Bohr, 597
    – de Broglie picture, 480
    -Sommerfeld quantization, 480, 521, 609

helium, 529, 543
Uetli Schwur, 608
Boltzmann, 23
equation, 422
boredom, 337, 579
Borges, J.L., 579
boundary orbits, 392
bounded operators, 682
Bowen, 26
brain, rat, 3, 27
branch cut, 357
branch point, 357
branch point singularity, 358
Burnett coefficient, 416

C_{3v} symmetry, 397
canonical
transformations, 82
canonical transformations, 612
Cartwright M.L., 604
Cauchy
criterion, 77
caustic, 492
ceiling function, 58, 144, 190
center, 79
chain rule, 87
matrix, 679
change
of coordinates, 41
chaos, 5, 6
caveats, 8
skeleton of, 9, 11
successes, 8
character
orthonormalitity, 662
representation, 658
characteristic
function, 98
polynomial, 245
chicken
heart palpitations, 5
circle map
critical, 428
coarse graining, 230
coarse-graining, 98
coding, *see* symbolic dynamics
collinear helium, 24
symbolic dynamics, 537
combinatorics
teaching, 212
complete
N-ary dynamics, 199
complexity
algorithmic, 258
confession
Kepler, 600
St. Augustine, 97
conjugacy, 41, 43
invariant, 91
smooth, 41, 53, 91, 697
topological, 213
conservation
equation, 617
phase-space volume, 83, 618
continuity
equation, 108
continuity equation, 490, 618
contour integral, 152
contracting
flow, 36, 40, 82
map, 132
stability eigenvalues, 89, 137
convergence
abscissa of, 309
mediocre, 674
radius, 147
super-exponential, 171
convexity, 130
coordinate
change, 41, 43
longitudinal, 501
transformations, 53
Copenhagen School, x, 609
correlation
decay
power law, 351
function, 186
spectrum, 187
time, 325
cost function, 279

covering
symbolic dynamics, 201
critical
point, 210
value, 211, 413
critical point, 88
cumulant
expansion, 244, 247, 298
Plemelj-Smithies, 686
curvature
correction, 295
expansion, 29, 295
cycle, 10
expansion, 17, 295
3-disk, 314
finite subshift, 304
stability ordered, 305
expon, 520
fundamental, 245, 295
limit, 126
point, 201
prime, 138, 201, 261, 269
3-disk, 283
Hénon map, 282
pruning, 251
stability, 87, 283
stable, 88
superstable, 88
unstable, 11
weigth, 150
cycles
fundamental, 672
cyclic
invariance, 269
symmetry, 242

D'Alambert's wave equation, 46
de Broglie wavelength, 484
decay
rate, 158
degree of freedom, 8
delta function
Dirac, 100, 483
density, 98, 617
evolution, 22
density of states
Green's function, 483
quantum, 483
semiclassical, 517
desymmetrization
3-disk, 403
determinant
Fredholm, 690
graph, 257
Hadamard, 148
spectral, 21, 148, 243
for flows, 149
trace relation, 244
trace-class operator, 683
deterministic
dynamics, 34
deterministic dynamics, 5
deterministic flow, 99
differential equations
ordinary
almost, 52
diffraction
Green's function, 562
Keller, 570
Sommerfeld, 570
diffusion
partial differential equations, 46
constant, 122
dike map, 213
dimension
box counting, 340
fractal, 339
generalized, 2
Hénon attractor, 284
information, 340
intrisic, 8
Dirac delta function, 100
Dirichlet series, 308
dissipative
map, 132
distribution, 47
divergence rate
local, 129
doubling map, 180
Duffing oscillator, 38

Duffing system, 50
dynamical
  system, 33, 34
    deterministic, 34
    gradient, 53
    smooth, 34
  systems
    equivalent, 42
  transitivity, 240
  zeta function, 16, 151
    Euler product rep., 151
dynamical system
  infinite, 55
  smooth, 17, 18, 29, 255, 466, 469, 471, 604
dynamics
  deterministic, 5
  hyperbolic, 205
  irreversible, 37
  reversible, 37
  stochastic, 5
  symbolic, 8, 198, 200, 328
  symmetry, 661
  topological, 198, 200, 202

edge, 203
eigendirections, 74
eigenfunction
  energy, 481
eigenfunctions
  Perron-Frobenius, 170
eigenstate, *see* eigenfunction
eigenvalue
  exponential spacing, 149
  spectral determinant, 171
  zero, 499
eigenvalues, 158
elliptic
  partial differential equations, 46
  stability, 83
enemy
  thy, 352
English
  plain, 201
ensemble
  microcanonical, 131
entropy
  barrier, 311
  topological, 6, 240, 252, 257
equation
  of variations, 74
equilibrium
  measure, *see* natural measure
  point, 39, 80, 102
equivalence
  of dynamical systems, 42
ergodic
  average, 104
  theorem
    multiplicative, 131
  theory, 104
escape
  intermittency, 361
  rate, 12, 321
    3-disk, 303, 315
escape rate, 102
  3-disk, 319
essential
  spectral radius, 176
essential spectral radius, 172
Euler
  formula, 171, 702
  limit, 81
  product, 155
  product rep.
    dynamical zeta function, 151
  totient function, 430
Euler-MacLaurin
  formula, 187
evolution
  group, 52, 696
  kernel
    probabilistic, 99
  operator, 20, 125
    quantum, 482
  semigroup, 126
expanding
  stability eigenvalues, 89, 137
expectation value, 121, 132
exponent

stability, 89
exponential
convergence, 147, 171
of a matrix, 76
exponential proliferation, 20, 258
extremal point, 498

false zeros, 155
Farey
map, 349, 375
mediant, 431
series, 430
tree, 431
Feynman, 597
Feynman path integral, 497
finite subshift
cycle expansion, 304
first return function, 58, 371
fixed point, 269
maps, 70
marginally stable, 349
Floquet multipliers, 136
flow, 34, 37
autonomous, 39, 275
contracting, 36, 40, 82
deterministic, 99
generator of, 106
Hamiltonian, 40, 64
hyperbolic, 83, 158
incompressible, 82, 618
infinite-dimensional, 45, 59
inverse hyperbolic, 83
linearized, 74
spectral determinant, 149
stability, 80
stationary, 39
stochastic, 99
stretch&fold, 211
flows
linear, 75
Fourier
mode, 46
spatial, 48
truncation, 49
fractal, 338
aggregates, 2, 328
dimension, 339
geometry of nature, 2
probabilistic, 2
science, 2
Fredholm
determinant, 690
integral equations, 174
integral operator, 173
Fredholm theory, 173
Fresnel integral, 498, 509
function
space
piecewise constant, 139
functional, 46
Lyapunov, 36
functions
L^2 square integrable, 172
analytic, 172
fundamental
cycle, 245
cycles, 672
domain, 206
collinear helium, 536
fundamental matrix, *see* Jacobian matrix

Galerkin truncation, 48
Gatto Nero
professor, 597
Gauss shift, 431
gaussian
integral, d-dimensional, 494
generating partition, 202
generator
of flow, 106
golden mean
pruning, 246
gradient algorithm, 280
grammar
symbolic dynamics, 202
grandmother
of fractals, 342
graph, 203
Markov, 203

Green's function
analogue of, 109
density of states, 483
diffraction, 562
energy dependent, 483
regularized, 518
semiclassical, 505, 511
short distance, 503
trace, 483
long orbits, 502
short distance, 503
group, 657
dynamical, 37
evolution, 52, 696
finite, 657
integration, 661
matrix, 658
order of, 657
representation, 658
semi-, 106
Gutzwiller
trace formula, 516

Hadamard determinant, 148
Hadamard product, 519
Hamilton
-Jacobi equation, 485
equations, 486
principal function, 488
Hamiltonian, 481, 487
flow, 64
spectral determinant, 156
stability, 614
flows, stability, 82, 611
repeller, periodic orbits, 290
Hankel function, 503
Hausdorff dimension
Hénon attractor, 284
Heaviside function, 484
Heisenberg, 609
picture, 681
Heisenberg, W, 105
Helfand moments, 416
helium, 529, 609
collinear, 24, 40, 55, 69, 555
cycles, 289, 556
eigenenergies, 556
fundamental domain, 536
Poincar/'e section, 555
stabilities, 556
stability, 289
Hénon
-Heiles
symbolic dynamics, 400
attractor, 70, 104
Hausdorff dimension, 284
Lyapunov exponent, 284
topological entropy, 284
map, 62, 66, 216, 280, 290
fixed points, 70, 216
inverse, 216
prime cycles, 282, 290
stability, 87
symmetries, 63
Hénon, M., 66
heroes
unsung, xii
Hessian matrix, 82
Hilbert
-Schmidt condition, 174
-Schmidt operators, 682
space, 481
Hopf bifurcation, 425
horseshoe, 208, 215
complete, 218
hydrodynamical
interpretation of QM, 504
hyperbolic
partial differential equations, 46
flow, 158
flows, 83
non-, 23
hyperbolicity assumption, 14, 137

in/out nodes, 79
inadmissible symbol sequence, 202
incompressible flow, 82
indecomposability, 240
metric, 199
index

Maslov, *see* topological index
induced map, 366
inertial manifold, 47, 48, 50
infinite-dimensional flows, 45, 59
inflection point, 426
information
dimension, 340
initial
conditions
sensitivity to, 5
point x_0, 14, 35, 74
state x_0, 14, 35
integrated observable, 118
integration
group, 661
Runge-Kutta, 54
intermittency, 67, 283, 348
escape rate, 361
piecewise linear model, 352
resummation, 368
stability ordering, 307
invariance
cyclic, 269
of flows, 90
symplectic, 82, 611
invariant
measure, 102
metric, 87, 91
topological, 87
inverse
hyperbolic flows, 83
inverse iteration, 270
Hamiltonian repeller, 290
inward/outward spirals, 79
irreducible
segment, 382
irreversibility, 22, 230
irreversible
dynamics, 37
iteration, 34
inverse, 270
Hamiltonian repeller, 290
itinerary, 9, 11, 198
bi-infinite, 200
future, 200, 212
past, 200

Jacobian, 81, 99
matrix, 14, 74, 621
mapping, 86
Jonquière function, 354, 375
Jordan form, 78

KAM
tori, 349
Keller
diffraction, 570
Keplerian orbit, 480
kernel
resolving, 175
kneading
determinant, 228
sequence, 213
theory, 213
value, 213
Koopman operator, 105, 109
Kramers, 609
KS, *see* Kustaanheimo-Stiefel
Kuramoto, Y., 50
Kuramoto-Sivashinsky system, 45
kurtosis, 133, 416
Kustaanheimo-Stiefel transformation, 532

L^2 function space, 172
Lagrangian, 488
manifold, 490
laminar states, 348
Laplace
transform, 21, 107, 141, 146, 248, 483, 505
transform, discrete, 138, 243
Laplacian, 45
Legendre transform, 489
Leibniz, 4
libration orbit, 539, *see* self–retracing
lifetime, 12
limit
cycle, 126
linear
flows, 75

stability, 73
linearized flow, 74
link, 203
Liouville
equation, 108
operator, 109
theorem, 83, 618
Littlewood J.E., 604
local
divergence rate, 129
stability, 73
logistic map, *see* unimodal
longitudinal
coordinate, 501
loop
intersecting, 245
Lorentz gas, 348
Lorenz, E.N., 66
loxodromic, 83
loxodromic eigenvalues, 614
Lozi map, 62, 66
Lyapunov
exponent, 6, 127
natural measure, 129
numerically, 128
functional, 36
time, 6, 8, 37
Lyapunov exponent
Hénon attractor, 284
numnerical, 131

,
$pS123$
Mandelbrot set, 328
manifold
stable, 91
map, 35
area preserving, 64
contracting, 132
dike, 213
dissipative, 132
expanding, 200
Hénon, 62, 280
prime cycles, 282
logistic, *see* unimodal
Lozi, 62, 66
once-folding, 215
order preserving, 213
orientation preserving, 64
orientation reversing, 64
quadratic, 63, 210
return, 11, 14, 57, 216
sawtooth, 383
stability, 86
tent, 210
Ulam, 315
unimodal, 210
maps
fixed point, 70
marginal
stability, 14, 349
fixed point, 349
stability eigenvalues, 89, 137
Markov
graph, 203
infinite, 247
partition, 417
finite, 200, 205
infinite, 229
not unique, 205
Maslov index, *see* topological index
material invariant, 617
matrix
exponential, 76
exponential of, 619
group, 658
normal, 78
of variations, 74
measure, 98
cylinder, 328
equipartition, 328
invariant, 102
natural, 103, 121
smooth, 119
mechanics
quantum, 480
statistical, 22
mediocre
convergence, 674
memory

finite, 223
metric
indecomposability, 199, 676
invariant, 87, 91
transitivity, 676
microcanonical ensemble, 131
Mira, C., 66
Misiurewicz, M., 66
mixing, 6, 7, 14, 104
mode
Fourier, 48
Moebius inversion, 250
monodromy matrix, 613
monodromy matrix, 136
Morse index, *see* topological index
mother
of fractals, 342
multifractals, 342, 462
multiplicative ergodic theorem, 131
multipoint shooting method, 272

natural measure, 103, 121, 129
nature
geometry of, 2
Nero
Gatto, professor, 597
neutral, *see* marginal
Newton's method, 271
convergence, 273
flows, 275
optimal surface of section, 278
node, 203
non-hyperbolicity, 90
non-wandering set, 36, 217
nonequilibrium, 407
norm, 77
normal
matrix, 78

obscure
foundations, 609
jargon, 198
topology, 49, 60
observable, 103, 118
ODE, *see* ordinary differential equations
open systems, 123
operator
evolution, 125
Hilbert-Schmidt, 682
Koopman, 105, 109
Liouville, 109
norm, 76
Perron-Frobenius, 100, 130
positive, 682
regularization, 689
resolvent, 107
semigroup
bounded, 107
shift, 201
trace-class, 681
orbit, 35
inadmissible, 213
Keplerian, 480
periodic, 201, 514, 516
returning, 514
order preserving map, 213
ordering
spatial, 215
ordinary differential equations
almost, 52
orientation
preserving map, 64
reversing map, 64
Oseledec multiplicative ergodic theorem, 131

palpitations
chicken hearts, 5
parabolic
partial differential equations, 46
paradise
this side of, 319
partial differential equations, 45
partition, 198, 202
generating, 202
infinite, 254, 257
Markov, 200
phase space, 98
partition function, 130
past topological coordinate, 220

PDE, *see* partial differential equations
period
  doubling
    repeller, 328
  relation to action, 541
periodic
  orbit, 201, 514, 516
  orbit condition, 269
  orbit extraction, 269–283
    Hamiltonian repeller, 290
    inverse iteration, 270
    multipoint shooting, 272
    Newton's method, 271–273
    relaxation algorithm, 280
  point, 10, 20, 201
    admissible, 243
    count, 249
    unstable, 11
Perron-Frobenius
  matrix, 240
  operator, 100, 130, 170
  theorem, 185, 191, 465
phase
  space
    partition, 98
    volume ,
      $pS123$
phase space, 34
  3-disk, 676
Piecewise
  linear maps, 375
piecewise
  constant functions, 139
piecewise linear model
  intermittency, 352
pinball, 4, *see* 3-disk
plain English, 201
Plemelj-Smithies cumulants, 686
Poincaré
  -Cartan theorem, 490
  return map, 57
    polynomial, 61
  section, 10, 11, 57, 65, 216
    3-disk, 65
point
  non-wandering, 36
  periodic, 10, 201
  wandering, 35
point scatterers, 574
Poisson
  bracket, 108, 109
  resummation, 21
  summation, 369
Polylogarithm, 354
polynomial
  characteristic, 245
  topological, 248
Pomeau, Y., 66
positive operators, 682
potential
  partial differential equations, 46
power law
  correlation decay, 351
pressure, 130
  thermodynamic, 130
prime cycle, 138, 201, 261, 269
  3-disk, 204, 261, 283
  count, 250
  Hénon map, 282
primitive cycle, *see* prime cycle
propagator, 482
  semiclassical, 493
  short time, 494, 503
  Van Vleck, 496
pruning, 9, 349
  front, 221
  golden mean, 246
  individual cycles, 251
  primary interval, 214
  symbolic dynamics, 202
pruning block, 202
pseudocycles, 294

quadratic map, 63
quantization
  Bohr-Sommerfeld, 480
  semiclassical, 513
quantum
  chaology, 26
  chaos, 27

evolution, 482
interference, 484
mechanics, 480
potential, 504
propagator, 482
resonances, 23
theory, old, 609

radius
of convergence, 147
random matrix theory, 26
recoding, 203, 206
rectangle, 218
rectification
flows, 43
recurrence, 36, 198
time, *see* return time
regularization, 519, 532
Green's function, 518
operator, 689
relaxation algorithm, 280
renormalization, 67
repeller, 24, 36, 123
single fixed point, 170
representation
character, 658
equivalent, 659
faithful, 659
matrix, 658
regular, 659
representative point, 34
residue
stability, 83
resolvent operator, 107
resolving kernel, 175
resonances
complex, 23
quantum, 23
resummation
intermittency, 368
return map, 11, 14, 216
return time, 372
distribution, 373
returning orbit, 514
reversible
dynamics, 37
Riemann zeta function, 308
Rössler system, 39, 208, 210
Ruelle, 26
zeta function, *see* dynamical zeta function
Runge-Kutta integration, 54
running orbits
Lorentz gas, 411
Rutherford, 529
Rydberg series, 543

S gradient
system, 53
saddle point, *see* stationary phase
saddle-node
bifurcation, 66
sawtooth map, 383
SBR measure, *see* natural measure
scatterer
point, 574
scattering
3-dimensional spheres, 92
Schrödinger
equation, 480
picture, 681
Schwartz, 47
section
Poincaré, 11, 57, 65
Selberg
product
eigenvalue index, 171
self-retracing cycle, 539
self-similar, 20
semiclassical
spectral determinant
collinear helium, 546
approximation, 485
density of states, 517
Green's function, 505, 511
propagator, 493
quantization, 513
wave function, 491
semiclassical zeta function, 520
semigroup, 106

dynamical, 37
evolution, 126
operator, 107
sensitivity to initial conditions, 5, 126
set
non-wandering, 217
shadowing, 17, 255
3-disk, 303
shift, 201
Bernoulli, 175
finite type, 202
full, 201
map, 426
operator, 201
sub-, 202
Sinai, 26
Sinai-Bowen-Ruelle measure, *see* natural measure
single fixed point
repeller, 170
spectral determinant, 170
singularity
branch point, 358
Sivashinsky, G.I., 50
skeleton of chaos, 9, 11
Smale, 8, 26, 160, 215, 228
wild idea, 149, 159
Smale S., 257, 604
smooth
conjugacy, 41, 53, 91, 697
dynamical system, 55
dynamics, 17, 18, 29, 34, 255, 466, 469, 471, 604
approximated, 671
Spectral determinant, 190
interaction, 472
measure, 119
potential, 67
Sommerfeld
diffraction, 570
space
analytic functions, 172
average, 103
averaging, 120
density functions, 139
spectral
determinant, 21, 148, 243
1-*d* maps, 155
2-*d* hyperbolic Hamiltonian flow, 156
entire, 149, 171
for flows, 149
infinite product rep., 150
single fixed point, 170
weighted, 156
spectral determinant
1-dimensional, 521
2-dimensional, 522
radius, 176
essential, 172
staircase, 484
spectral stability, 419
spectrum
Balmer, 480
St. Augustine, 97
stability
billiards, 83
cycle, 283
eigenvalues, 89, 137
exponents, 89
flow, 80
Hamiltonian flow, 614
Hamiltonian flows, 82, 611
linear, 73
maps, 86
neutral, *see* marginal
ordering
cycle expansions, 305
intermittent flows, 307
residue, 83
spectral, 419
structural, 218, 221, 256, 419
window, 90
stable
cycle, 88
manifold, 14, 91, 216
flow, 91
map, 91
stagnation point, 39, *see* equilibrium point

staircase
mean eigenvalue density, 546
spectral, 484
standing orbit
Lorentz gas, 411
state, 198, 203
set, 198
state space, 34, *see* phase space
stationary
flow, 39
phase approximation, 496, 498, 509, 515, 561, 573, 717
point, 39, *see* equilibrium point
state, 102
statistical mechanics, 22
Sterling formula, 509
stochastic dynamics, 5
stochastic flow, 99
stochasticity assumption, 372
Stokes theorem, 490
stosszahlansatz, 23, 422
strange
attractor, 36, 40
strange attractor, 126
stretch & fold dynamics, 62
stretch&fold flow, 211
strobe method, 57
structural stability, 218, 221, 256, 419
subshift, 202
finite type, 203, 204, 221, 224
superstable cycle, 88
surface of section
optimal, 278
survival probability, 13, *see* escape rate
symbol
sequence
inadmissible, 202
square, 218
symbolic
dynamics, 328
at a bifurcation, 67
complete N-ary, 199
covering, 201
symbolic dynamics, 8, 198, 200
3-disk, 32, 204, 233
binary
collinear helium, 537
coding, 202
Markov graph, 304
complete, 211, 218
grammar, 202
Hénon-Heiles, 400
pruned, 202
recoding, 203, 206
unimodal, 211
symmetries
Hénon map, 63
symmetry
C_{3v}, 397
3-disk, 206, 382, 397, 403
cyclic, 242
discrete, 205
dynamical system, 661
symplectic
form, 82
group $Sp(2D)$, 612
integrator, 620
invariance, 82, 611
transformation, 82, 109, 228
systems
open, 123

tangent
bundle, 38
tangent space, 74
Tauberian theorem, 376
teaching
combinatorics, 212
tessalation
smooth dynamics, 671
thermodynamical
pressure, 130
3-body problem, 24, 513, 529, 599, 610
3-dimensional sphere
scattering, 92
3-disk
boundary orbits, 392
convergence, 178, 671

cycle
analytically, 95
count, 400, 641
expansion, 314
escape rate, 167, 303, 315, 319
fractal dimension, 338
geometry, 65
hyperbolicity, 137
phase space, 11, 338, 676
pinball, 4, 65–67
point scatterer, 574
prime cycles, 16, 204, 261, 283
shadowing, 303
simulator, 71, 94, 96
symbolic dynamics, 9, 32, 204, 233
symmetry, 206, 382, 397, 403
transition matrix, 199
time
arrow of, 22
as parametrization, 42
average, 103, 128
averaging, 118
ceiling function, *see* ceiling function
ordered integration, 80
topological
conjugacy, 213
dynamics, 198, 200, 202, 203
entropy, 6, 240, 252
Hénon attractor, 284
equivalence, 79
future coordinate, 212
index, 492
invariant, 87
parameter, 214
polynomial, 248
trace formula, 243
transitivity, 240
zeta function, 247, 248
topological index, 517, 610
totient function, 430
trace
formula
classical, 21
flows, 142
Gutzwiller, 516
maps, 138
topological, 243, 248
local, 242
trace-class operator, 681
determinant, 683
trajectory, 35, 75, *see* orbit
transfer
matrix, 101
transfer operator, 159
spectrum, 148
transformation
canonical, 228
coordinate, 53
symplectic, 228
transient, 35, 82, 199, 227, 286
transition matrix, 199, 240, 242
3-disk, 199
transverse
stability, 502
Trotter product formula, 681
truncation
Galerkin, 48
truncations
Fourier, 49

Ulam map, 288, 315
unimodal map, 210
symbolic dynamics, 211
unstable
cycle, 11
manifold, 14, 91, 216
map, 91
manifold, flow, 91
periodic point, 11
unsung
heroes, xii

Van Vleck
propagator, 496
vector field, 37
vector fields
singularities, 43
vertex, 203
visitation sequence, *see* itinerary

Viswanath, 597
volume preservation, 85
von Neumann
    ergodicity, 109

wandering point, 35
wave
    partial differential equations, 46
wave function
    semiclassical, 491
    WKB, 492
weight
    multiplicative, 29
Wentzel-Kramers-Brillouin, 485, *see* WKB
winding number, 426, 428
WKB, 47, 485
    wave function, 492

Young, L.-S., 66

zero eigenvalue, 499
zeros
    false, 155
zeta function
    Artin-Mazur, 247
    dynamical, 16, 151
    Ruelle, *see* dynamical
    topological, 247, 248

Part II

Classical and Quantum Chaos: Material available on www.nbi.dk/ChaosBook/

---

version 9.2.3 **Feb 26 2002**

printed June 19, 2002

www.nbi.dk/ChaosBook/ **comments to: predrag@nbi.dk**

Appendix A

What reviewers say

A.1 N. Bohr

"The most important work since that Schrödinger killed the cat."

A.2 R.P. Feynman

"Great doorstop!"

A.3 Divakar Viswanath

I have read the first six chapters. The introduction chapter is fabulous. The computation of the escape rate for the pin-ball game in the first chapter is absolutely riveting. There can't be a better way to begin. I find the book to be very stimulating.

A.4 Professor Gatto Nero

This book, which I have received unsolicited from the Szczsyrk Oblast Agricultural and Mazuth Office Press appears to be a collage of LaTeX clips from random papers authored by the motley collection of co-authors whose number one usually associates with an experimental high energy *Phys. Rev. Letter*, rather than a book that aspires to be the Landau-Lifshitz of chaos.

Entire rain forests went down so this not inconsiderable tome can be printed and bound. Why these ravings were not left on the Web where they more properly belong is not explained anywhere in the text. If it is an antiBourbaki, as one has in the antimatter world, then why not do the real thing? A Landau-Lifshitz for nonlinear Science, written as it should be done. The nonScience book to end all nonScience books.

Appendix B

A brief history of chaos

Laws of attribution

1. **Arnold's Law**: everything that is discovered is named after someone else (including Arnold'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

B.1 Chaos is born

(R. Mainieri)

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.

B.1.1 Three-body problem

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.

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 domain 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 phase 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.

B.1.2 Ergodic hypothesis

The second problem that played a key role in development of chaotic dynamics was the ergodic hypothesis of Boltzmann. Maxwell and Boltzmann had combined the mechanics of Newton with notions of probability in order to create statistical mechanics, deriving thermodynamics from the equations of mechanics. 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 conservations law 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 our own 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.

B.1.3 Nonlinear oscillators

The third problem that was very influential in the development of the theory of chaotic dynamical systems was the work on 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.

B.2 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 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, 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 Smale.

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 phase 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.

B.3 Chaos with us

(R. Mainieri)

In the fall of 1961 Steven Smale was invited to Kiev where he met Arnold, 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.

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 phase 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 phase 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 one-dimensional maps. The topology of the orbits for parabola-like maps was worked out in 1973 by Metropolis, Stein, and Stein. The more general one-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 one-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 sys-

tems, 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.

Remark **B.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 nonwandering 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 **B.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 ϵ^n, or faster.

Ya.G. Sinai, L. Bunimovich and 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 ϵ neighborhood around it. Given that the Lebesgue measure is ϵ^α and the stability grows not faster than (distance)^n, A. Katok and J.-M. Strelcyn prove 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 discontinutities, and given that the stability grows not faster than (distance)^n, Ya.B. Pesin proved that the Lyapunov exponent is non-zero, and that SBR measure exists. He also proved that the Lorenz, Lozi and Byelikh attractors satisfy these conditions.

In the the systems were uniformly hyperbolic, all trouble was 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 excizing the regions that do not separate. Hence there are 3 levels of ergodic systems:

1. Anosov flow
2. Anosov flow + singularity set
   - the Hamiltonian systems: general case A. Katok and J.-M. Strelcyn, billiards Ya.G. Sinai and L. Bunimovich.
   - the dissipative case: Ya.B. Pesin
3. Hénon
   - The first proof was given by M. Benedicks and L. Carleson [**?**].
   - A more readable proof is given in M. Benedicks and L.-S. Young [**?**]

(based on Ya.B. Pesin's comments)

B.3.1 Periodic orbit theory

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, Ruelle, and many others; 2. *quantum theory* initiated by Bohr, with the modern "chaotic" formulation by Gutzwiller; and 3. analytic *number theory* initiated by Riemann and formulated as a spectral problem by Selberg. 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, that is, 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.

B.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 added the topological indices in 1980, and in 1991 Wintgen and collaborators 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.

Remark **B.3** Sources. This tale, aside from a few personal recollections, is in large part lifted from Abraham Pais' accounts of the demise of the old quantum theory [**?**, **?**], as well as Jammer's account [1]. The helium spectrum is taken up in chapter 23. In August 1994 Dieter Wintgen died in a climbing accident in the Swiss Alps.

Appendix C

Stability of Hamiltonian flows

C.1 Symplectic invariance

(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. 2.2.2. We assume you are at home with Hamiltonian formalism. If you would like to see the Hamilton's equations derived, Hamilton's original line of reasoning is retraced in sect. 21.2.1. The evolution equations for any p, q dependent quantity $Q = Q(p, q)$ are given by (5.31).

☞ sect. 21.2.1

In terms of the Poisson brackets, the time evolution equation for $Q = Q(p, q)$ is given by (5.33). We now recast the symplectic condition (4.37) in a form convenient for using the symplectic constraints on $\mathbf{J}$. Writing $x(t) = x' = [p', q']$ and the Jacobian matrix and its inverse

$$\mathbf{J} = \begin{pmatrix} \frac{\partial q'}{\partial q} & \frac{\partial q'}{\partial p} \\ \frac{\partial p'}{\partial q} & \frac{\partial p'}{\partial p} \end{pmatrix}, \qquad \mathbf{J}^{-1} = \begin{pmatrix} \frac{\partial q}{\partial q'} & \frac{\partial q}{\partial p'} \\ \frac{\partial p}{\partial q'} & \frac{\partial p}{\partial p'} \end{pmatrix}, \tag{C.1}$$

we can spell out the symplectic invariance condition (4.37):

$$\begin{aligned} \frac{\partial q'_k}{\partial q_i}\frac{\partial p'_k}{\partial q_j} - \frac{\partial p'_k}{\partial q_i}\frac{\partial q'_k}{\partial q_j} &= 0 \\ \frac{\partial q'_k}{\partial p_i}\frac{\partial p'_k}{\partial p_j} - \frac{\partial p'_k}{\partial p_i}\frac{\partial q'_k}{\partial p_j} &= 0 \\ \frac{\partial q'_k}{\partial q_i}\frac{\partial p'_k}{\partial p_j} - \frac{\partial p'_k}{\partial q_i}\frac{\partial q'_k}{\partial p_j} &= \delta_{ij}\,. \end{aligned} \tag{C.2}$$

From (4.38) we obtain

$$\frac{\partial q_i}{\partial q'_j} = \frac{\partial p'_j}{\partial p_i}\,, \quad \frac{\partial p_i}{\partial p'_j} = \frac{\partial q'_j}{\partial q_i}\,, \quad \frac{\partial q_i}{\partial p'_j} = -\frac{\partial q'_j}{\partial p_i}\,, \quad \frac{\partial p_i}{\partial q'_j} = -\frac{\partial p'_j}{\partial q_i}\,. \tag{C.3}$$

Taken together, (C.3) and (C.2) imply that the flow conserves the $[p,q]$ Poisson brackets

$$\begin{aligned} [q_i,q_j] &= \frac{\partial q_i}{\partial p'_k}\frac{\partial q_j}{\partial q'_k} - \frac{\partial q_j}{\partial p'_k}\frac{\partial q_i}{\partial q'_k} = 0 \\ [p_i,p_j] &= 0\,, \qquad [p_i,q_j] = \delta_{ij}\,, \end{aligned} \tag{C.4}$$

that is, 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 $\mathbf{J}$ are linearly independent, as it behooves group elements of the symplectic group $Sp(2D)$.

We have now succeeded in making the full set of constraints explicit - as we shall see in appendix D, this will enable us to implement dynamics in such a way that the symplectic invariance will be automatically preserved.

Consider the symplectic product of two infinitesimal vectors

$$\begin{aligned} (\delta x, \delta\hat{x}) &= \delta x^T \omega \delta\hat{x} = \delta p_i \delta\hat{q}_i - \delta q_i \delta\hat{p}_i \\ &= \sum_{i=1}^{D} \{\text{oriented area in the } (p_i,q_i) \text{ plane}\}\,. \end{aligned} \tag{C.5}$$

Time t later we have

$$(\delta x', \delta\hat{x}') = \delta x^T \mathbf{J}^T \omega \mathbf{J} \delta\hat{x} = \delta x^T \omega \delta\hat{x}\,.$$

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 δx 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 symplectic area spanned by the three final points is the same as that spanned by the inital 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 (p_i, q_i) plane, are separately conserved:

$$\int_{\mathcal{V}} dp \wedge dq = \oint_{\Omega\mathcal{V}} p \cdot dq = \text{invariant}\,. \tag{C.6}$$

C.2 Monodromy matrix for Hamiltonian flows

(G. Tanner)

It is not the Jacobian 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 Jacobian matrix $\mathbf{J}$. An initial displacement in the direction of the flow $x = \omega\nabla H(x)$ transfers according to $\delta x(t) = x_t(t)\delta t$ with δt time independent. The projection of any displacement on δx on $\nabla H(x)$ is constant, i.e. $\nabla H(x(t))\delta x(t) = \delta E$. We get the equations of motion for the monodromy matrix directly choosing a suitable local coordinate system on the orbit $x(t)$ in form of the (non singular) transformation $\mathbf{U}(x(t))$:

$$\tilde{\mathbf{J}}(x(t)) = \mathbf{U}^{-1}(x(t))\,\mathbf{J}(x(t))\,\mathbf{U}(x(0)) \tag{C.7}$$

These lead to

$$\begin{aligned} & \dot{\tilde{\mathbf{J}}} &=& \tilde{\mathbf{L}}\,\tilde{\mathbf{J}} \\ \text{with} \qquad & \tilde{\mathbf{L}} &=& \mathbf{U}^{-1}(\mathbf{L}\mathbf{U} - \dot{\mathbf{U}}) \end{aligned} \tag{C.8}$$

Note that the properties a) – c) are only fulfilled for $\tilde{\mathbf{J}}$ and $\tilde{\mathbf{L}}$, if $\mathbf{U}$ itself is symplectic.
Choosing $x_E = \nabla H(t)/|\nabla H(t)|^2$ and x_t as local coordinates uncovers the two trivial eigenvalues 1 of the transformed matrix in (C.7) at any time t. Setting $\mathbf{U} = (x_t^T, x_E^T, x_1^T, \ldots, x_{2d-2}^T)$ gives

$$\tilde{\mathbf{J}} = \begin{pmatrix} 1 & * & * & \ldots & * \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & * & & & \\ \vdots & \vdots & & \mathbf{m} & \\ 0 & * & & & \end{pmatrix}; \qquad \tilde{\mathbf{L}} = \begin{pmatrix} 0 & * & * & \ldots & * \\ 0 & 0 & 0 & \ldots & 0 \\ 0 & * & & & \\ \vdots & \vdots & & \mathbf{l} & \\ 0 & * & & & \end{pmatrix}, \tag{C.9}$$

The matrix $\mathbf{m}$ is now the monodromy matrix and the equation of motion are given by

$$\dot{\mathbf{m}} = \mathbf{l}\,\mathbf{m}. \tag{C.10}$$

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.

$$\mathbf{U}(t) = (x_t, x_1, x_E, x_2) = \begin{pmatrix} \dot{x} & -\dot{y} & -\dot{u}/q^2 & -\dot{v}/q^2 \\ \dot{y} & \dot{x} & -\dot{v}/q^2 & \dot{u}/q^2 \\ \dot{u} & \dot{v} & \dot{x}/q^2 & -\dot{y}/q^2 \\ \dot{v} & -\dot{u} & \dot{y}/q^2 & \dot{x}/q^2 \end{pmatrix} \tag{C.11}$$

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 (C.13). 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 Ψ 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

$$\lambda = \frac{\mathrm{Tr}(\mathbf{m}) \pm \sqrt{\mathrm{Tr}(\mathbf{m})^2 - 4}}{2}, \tag{C.12}$$

i.e. $\mathrm{Tr}(\mathbf{m}) = 2$ separates stable and unstable behavior.

The $\mathbf{l}$ matrix elements for the local transformation (C.11) are

$$\tilde{\mathbf{l}}_{11} = \frac{1}{q}[(h_x^2 - h_y^2 - h_u^2 + h_v^2)(h_{xu} - h_{yv}) + 2(h_x h_y - h_u h_v)(h_{xv} + h_{yu})$$

$$
\begin{aligned}
& -(h_x h_u + h_y h_v)(h_{xx} + h_{yy} - h_{uu} - h_{vv})] \\
\tilde{\mathbf{I}}_{12} &= \frac{1}{q^2}[(h_x^2 + h_v^2)(h_{yy} + h_{uu}) + (h_y^2 + h_u^2)(h_{xx} + h_{vv}) \\
& -2(h_x h_u + h_y h_v)(h_{xu} + h_{yv}) - 2(h_x h_y - h_u h_v)(h_{xy} - h_{uv})] \\
\tilde{\mathbf{I}}_{21} &= -(h_x^2 + h_y^2)(h_{uu} + h_{vv}) - (h_u^2 + h_v^2)(h_{xx} + h_{yy}) \\
& +2(h_x h_u - h_y h_v)(h_{xu} - h_{yv}) + 2(h_x h_v + h_y h_u)(h_{xv} + h_{yu}) \\
\tilde{\mathbf{I}}_{22} &= -\tilde{\mathbf{I}}_{11}, \qquad \text{(C.13)}
\end{aligned}
$$

with h_i, h_{ij} is the derivative of the Hamiltonian H with respect to the phase space coordinates and $q = |\nabla H|^2$.

Remark **C.1** The sign convention of the Poisson bracket. The Poisson bracket is antisymmetric in its arguments and there is a freedom to define it with the opposite sign convention. When such freedoms exist in physics, it is certain that both conventions are in use and this is no exception. In several texts you will see the right hand side of (5.32) defined as $[B, A]$ so that (5.33) is $\frac{dQ}{dt} = [Q, H]$. As long as one is consistent, there should be no problem.

Remark **C.2** The sign convention of ω. The overall sign of ω, the symplectic invariant in (4.33), 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.

Appendix D

Implementing evolution

The large body of accrued wisdom on the subject of flows called fluid dynamics is about physicaly real flows of media with (typically) continuous densities. On the other hand, the flows in state spaces of complicated systems are not only abstract, but may lead to fractal or other forms of complication in the densities. The latter frequently require more abstract tools that we develop in this text. Before sketching those, it might be helpful to outline the more tangible fluid dynamical vision.

D.1 Material invariants

> I'm a material girl in a material world.
> Madonna, *Material Girl*

(E.A. Spiegel and P. Cvitanović)

We consider first the simplest property of a flow called a *material invariant* by fluid dynamicists, who have no choice but to ponder such matters. 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 the fluid is green. As $I(x) = I(f^t(x))$ 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

$$\partial_t I + v_i \partial_i I = 0\,. \tag{D.1}$$

Let the *density* of representative points be $\rho(x,t)$. To each representative point x we assign the value $I(x)$ of some physical property of a system. 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

$$\partial_t \int_{\mathcal{V}} dx\, \rho I = -\int_{\partial\mathcal{V}} d\sigma\, n_i v_i \rho I\,, \tag{D.2}$$

where $\mathcal{V}$ is an arbitrary volume in $\mathcal{M}$, $\partial\mathcal{V}$ is its surface and $\mathbf{n}$ is its outward normal. Repeated indices are summed over throughout and ∂_t is the partial derivative with respect to time.

We may use the divergence theorem to turn the surface integral into a volume integral and obtain

$$\int_{\mathcal{V}} \left[\partial_t(\rho I) + \partial_i(v_i \rho I)\right] dx = 0\,,$$

where ∂_i is the partial derivative operator with respect to x_i. Since the integration is over an arbitrary volume, we conclude that

$$\partial_t(\rho I) + \partial_i(\rho I v_i) = 0\,. \tag{D.3}$$

The choice $I \equiv 1$ yields the *continuity equation* for the density:

$$\partial_t \rho + \partial_i(\rho v_i) = 0\,. \tag{D.4}$$

If the density itself is a material invariant, combining (D.1) and (D.4) we conclude that $\partial_i v_i = 0$ and $\mathbf{J}^t(x_0) = 1$. An example of such incompressible flow is the Hamiltonian flow of sect. 4.4. For incompressible flows the continuity equation (D.4) becomes a statement of conservation of the phase-space volume (see sect. 4.4), or the Liouville theorem

$$\partial_t \rho + v_i \partial_i \rho = 0\,. \tag{D.5}$$

D.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 (5.25) as

$$\mathcal{A} = \left.\frac{d}{dt}\mathcal{K}^t\right|_{t=0} .$$

If we now take the derivative at arbitrary times we get

$$\begin{aligned}\left(\frac{d}{dt}\mathcal{K}^t\psi\right)(x) &= \lim_{\eta\to 0}\frac{\psi(f^{t+\eta}(x)) - \psi(f^t(x))}{\eta} \\ &= v_i(f^t(x))\left.\frac{\partial}{\partial\tilde{x}_i}\psi(\tilde{x})\right|_{\tilde{x}=f^t(x)} \\ &= \left(\mathcal{K}^t\mathcal{A}\psi\right)(x)\end{aligned}$$

which can be formally integrated like an ordinary differential equation yielding 5.11 on p. 115

$$\mathcal{K}^t = e^{t\mathcal{A}} \tag{D.6}$$

which guarantees that the Laplace transform manipulations in sect. 5.4 are correct. Though the formal expression of the semigroup (D.6) is quite simple one has to take care in implementing its action. If we express the exponential through the power series

$$\mathcal{K}^t = \sum_{k=0}^{\infty}\frac{t^k}{k!}\mathcal{A}^k , \tag{D.7}$$

dealing with powers of $\mathcal{A}$ we encounter the problem that the infinitesimal generator (5.25) contains non-commuting pieces, that is, 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

$$\mathcal{K}^t = \prod_{m=1}^{t/\tau}\mathcal{K}^\tau$$

as the starting point for a discretized approximation (with time step τ) to the continuous time dynamics. Omitting terms from the second order onwards in the expansion of $\mathcal{K}^\tau$ yields an error of order $O(\tau^2)$. This might be acceptable if the time step τ is sufficiently small: even in this case one has to be careful, since

$$\left(\mathcal{A}\mathcal{K}^s\psi\right)(x) \;=\; v_i(f^s(x))\,\frac{\partial\psi}{\partial\tilde{x}_i}\bigg|_{\tilde{x}=f^s(x)}$$

In practice we write the Euler product

$$\mathcal{K}^t \;=\; \prod_{m=1}^{t/\tau}\left(1+\tau\mathcal{A}_{(m)}\right) \;+\; O(\tau^2) \tag{D.8}$$

where

$$\left(\mathcal{A}_{(m)}\psi\right)(x) \;=\; v_i(f^{m\tau}(x))\,\frac{\partial\psi}{\partial\tilde{x}_i}\bigg|_{\tilde{x}=f^{m\tau}(x)}$$

As far as the x dependence is concerned, $e^{\delta\tau\mathcal{A}_i}$ acts as

$$e^{\delta\tau\mathcal{A}_i}\begin{Bmatrix} x_1 \\ \cdot \\ x_i \\ x_d \end{Bmatrix} \to \begin{Bmatrix} x_1 \\ \cdot \\ x_i+\delta\tau v_i(x) \\ x_d \end{Bmatrix}. \tag{D.9}$$

2.8 on p. 54

We see that the product form (D.8) 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.

D.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 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 (D.8) by spliting it into infinitesimal steps

D.2 on p. 623

and keeping terms up to $\delta\tau^2$ in

$$\mathcal{K}^{\delta\tau} = \hat{\mathcal{K}}^{\delta\tau} + \frac{1}{24}(\delta\tau)^3[\mathcal{B}+2\mathcal{C},[\mathcal{B},\mathcal{C}]] + \cdots \tag{D.10}$$

where

$$\hat{\mathcal{K}}^{\delta\tau} = e^{\frac{1}{2}\delta\tau\mathcal{B}} e^{\delta\tau\mathcal{C}} e^{\frac{1}{2}\delta\tau\mathcal{B}} \tag{D.11}$$

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 (5.28), a free flight by $\frac{1}{2}\delta\tau\mathcal{B}$, scattering by $\delta\tau\partial V(q')$, followed again by $\frac{1}{2}\delta\tau\mathcal{B}$ free flight:

$$\begin{aligned}
e^{\frac{1}{2}\delta\tau\mathcal{B}}\begin{Bmatrix} q \\ p \end{Bmatrix} &\to \begin{Bmatrix} q' \\ p' \end{Bmatrix} = \begin{Bmatrix} q - \frac{\delta\tau}{2}p \\ p \end{Bmatrix} \\
e^{\delta\tau\mathcal{C}}\begin{Bmatrix} q' \\ p' \end{Bmatrix} &\to \begin{Bmatrix} q'' \\ p'' \end{Bmatrix} = \begin{Bmatrix} q' \\ p' + \delta\tau\partial V(q') \end{Bmatrix} \\
e^{\frac{1}{2}\delta\tau\mathcal{B}}\begin{Bmatrix} q'' \\ p'' \end{Bmatrix} &\to \begin{Bmatrix} q''' \\ p''' \end{Bmatrix} = \begin{Bmatrix} q' - \frac{\delta\tau}{2}p'' \\ p'' \end{Bmatrix}
\end{aligned} \tag{D.12}$$

Collecting the terms we obtain an integration rule for this type of symplectic flow which is better than the straight Euler integration (D.9) as it is accurate up to order $\delta\tau^2$:

$$\begin{aligned}
q_{n+1} &= q_n - \delta\tau\, p_n - \frac{(\delta\tau)^2}{2}\partial V\left(q_n - \delta\tau p_n/2\right) \\
p_{n+1} &= p_n + \delta\tau\partial V\left(q_n - \delta\tau p_n/2\right)
\end{aligned} \tag{D.13}$$

The Jacobian matrix of one integration step is given by

$$\mathbf{J} = \begin{pmatrix} 1 & -\delta\tau/2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \delta\tau\partial V(q') & 1 \end{pmatrix} \begin{pmatrix} 1 & -\delta\tau/2 \\ 0 & 1 \end{pmatrix} . \tag{D.14}$$

Note that the billiard flow (4.49) is an example of such symplectic integrator. In that case the free flight is interupted by instantaneous wall reflections, and can be integrated out.

Remark **D.1** Symplectic integration. A good starting point on the symplectic integrators literature is represented by the reviews [3] and [4].

References

[D.1] B.A. Shadwick, J.C. Bowman, and P.J. Morrison, *Exactly Conservative Integrators*, chao-dyn/9507012, Submitted to *SIAM J. Sci. Comput.*

[D.2] D.J.D Earn, *Symplectic integration without roundoff error*, astro-ph/9408024.

[D.3] P.J. Channell and C. Scovel, *Nonlinearity* **3**, 231 (1990).

[D.4] J.M. Sanz-Serna and M.P. Calvo, *Numerical Hamiltonian problems* (Chapman and Hall, London, 1994).

Exercises

D.1 Symplectic volume preservation. Check that the sequence of mappings (D.12) is volume preserving, $\det \hat{U} = 1$.

D.2 Noncommutativity. Check that the commutators in (D.10) are not vanishing by showing that

$$[\mathcal{B}, \mathcal{C}] = -p \left(V'' \frac{\partial}{\partial p} - V' \frac{\partial}{\partial q} \right) .$$

D.3 Symplectic leapfrog integrator. Implement (D.13) for 2-dimensional Hamiltonian flows; compare it with Runge-Kutta integrator by integrating trajectories in some (chaotic) Hamiltonian flow.

Appendix E

Symbolic dynamics techniques

The kneading theory for unimodal mappings is developed in sect. E.1. Pruning theory for Bernoulli shifts (an exercise mostly of formal interest) is discussed in sect. **??**. The prime factorization for dynamical itineraries of sect. E.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.

E.1 Topological zeta functions for infinite subshifts

(Per Dahlqvist)

The Markov graph methods outlined in chapter **??** 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 $\ldots x_{i-1}x_ix_{i+1}\ldots$ onto a sequence of symbols $\ldots s_{i-1}s_is_{i+1}\ldots$ where

$$\begin{aligned} s_i &= 0 \quad x_i < x_c \\ s_i &= C \quad x_i = x_c \\ s_i &= 1 \quad x_i > x_c \end{aligned} \tag{E.1}$$

$I(C)$	$\zeta_{top}^{-1}(z)/(1-z)$
$1C$	
$101C$	
$1011101C$	
$H^\infty(1)$	$\prod_{n=0}^{\infty}(1-z^{2^n})$
$10111C$	
$1011111C$	
101^∞	$(1-2z^2)/(1+z)$
$10111111C$	
$101111C$	
$1011C$	
$101101C$	
$10C$	$(1-z-z^2)$
$10010C$	
$100101C$	

$I(C)$	$\zeta_{top}^{-1}(z)/(1-z)$
$1001C$	
$100111C$	
$10011C$	
$100110C$	
$100C$	
$100010C$	
$10001C$	
$100011C$	
$1000C$	
$100001C$	
$10000C$	
$100000C$	
10^∞	$(1-2z)/(1-z)$

Table E.1: All ordered kneading sequences up to length seven, as well as some longer kneading sequences. Harmonic extension $H^\infty(1)$ is defined below.

and x_c is the critical point of the map (that is maximum of g). In addition to the usual binary alphabet we have added a symbol C for the critical point. The kneading sequence K_Λ is the itinerary of the critical point. The crucial observation is that no periodic orbit can have a topological coordinate (see sect. E.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 Λ.

The kneading sequence can be of one of three types

1. It maps to the critical point again, after n iterations. If so, we adopt the convention to terminate the kneading sequence with a C, and refer to the kneading sequence as finite.
2. Preperiodic, that is it is infinite but with a periodic tail.
3. Aperiodic.

As an archetype unimodal map we will choose the *tent map*

$$x \mapsto f(x) = \begin{cases} \Lambda x & x \in [0, 1/2] \\ \Lambda(1-x) & x \in (1/2, 1] \end{cases} , \tag{E.2}$$

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

$$1/\zeta(z) = \prod_p \left(1 - \frac{z^{n_p}}{|\Lambda_p|}\right) = \prod_p \left(1 - \left(\frac{z}{\Lambda}\right)^{n_p}\right) = 1/\zeta_{\rm top}(z/\Lambda)$$

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* Λ. For general unimodal maps the corresponding statement is that the kneading sequence is aperiodic for almost all topological entropies.

For a given periodic kneading sequence of period n, $\underline{K}_\Lambda = PC = s_1 s_2 \dots s_{n-1} C$ there is a simple expansion for the topological zeta function. Then the expanded zeta function is a polynomial of degree n

$$1/\zeta_{\rm top}(z) = \prod_p (1 - z_p^n) = (1 - z) \sum_{i=0}^{n-1} a_i z^i \ , \qquad a_i = \prod_{j=1}^{i} (-1)^{s_j} \tag{E.3}$$

and $a_0 = 1$.

Aperiodic and preperiodic kneading sequences are accounted for by simply replacing n by ∞.

Example. Consider as an example the kneading sequence $K_\Lambda = 10C$. From (E.3) we get the topological zeta function $1/\zeta_{top}(z) = (1 - z)(1 - z - z^2)$, see table E.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 (E.3) we get the topological zeta function $1/\zeta_{top}(z) = (1 - z)(1 - 2z^2)/(1 + z)$, see table E.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 (E.3) is that the sequence $\{a_i\}$ has a periodic tail if and only if the kneading sequence has one (however, their period may differ by a factor of two). We know already that the kneading sequence is aperiodic for almost all Λ.

The analytic structure of the function represented by the infinite series $\sum a_i z_i$ with unity as radius of convergence, depends on whether the tail of $\{a_i\}$ is periodic or not. If the period of the tail is N we can write

$$1/\zeta_{top}(z) = p(z) + q(z)(1 + z^N + z^{2N} \ldots) = p(z) + \frac{q(z)}{1 - z^N} \quad ,$$

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 [**?**], 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 fig. 11.2. 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 (E.2), for almost all Λ.

Let us now focus on the relation between the analytic structure of the topological zeta function and the number of periodic orbits, or rather (11.5), the number N_n of fixed points of $f^n(x)$. The trace formula is (see sect. 11.4)

$$N_n = \mathrm{tr}\, T^n = \frac{1}{2\pi i} \oint_{\gamma_r} dz\, z^{-n} \frac{d}{dz} \log \zeta_{top}^{-1}$$

where γ_r is a (circular) contour encircling the origin $z = 0$ in clockwise direction. Residue calculus turns this into a sum over zeros z_0 and poles z_p of ζ_{top}^{-1}

$$N_n = \sum_{z_0 : r < |z_0| < R} z_0^{-n} - \sum_{z_p : r < |z_p| < R} z_0^{-n} + \frac{1}{2\pi i} \oint_{\gamma_R} dz\, z^{-n} \frac{d}{dz} \log \zeta_{top}^{-1}$$

and a contribution from a large circle γ_R. For meromorphic topological zeta functions one may let $R \to \infty$ with vanishing contribution from γ_R, and N_n will be a sum of exponentials.

The leading zero is associated with the topological entropy, as discussed in chapter **??**.

We have also seen that for preperiodic kneading there will be poles on the unit circle.

To appreciate the role of natural boundaries we will consider a (very) special example. Cascades of period doublings is a central concept for the description of unimodal maps. This motivates a close study of the function

$$\Xi(z) = \prod_{n=0}^{\infty}(1 - z^{2^n}) \quad . \tag{E.4}$$

This function will appear again when we derive (E.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 γ_R cannot be pushed away to infinity, R is restricted to $R \leq 1$ and N_n is entirely determined by $\int_{\gamma_R}$ which picks up its contribution from the natural boundary.

We have so far studied the analytic structure for some special cases and we know that the unit circle is a natural boundary for almost all Λ. But how does it look out there in the complex plane for some typical parameter values? To explore that we will imagine a journey from the origin $z = 0$ out towards the unit circle. While traveling we let the parameter Λ change slowly. The trip will have a distinct science fiction flavor. The first zero we encounter is the one connected to the topological entropy. Obviously it moves smoothly and slowly. When we move outward to the unit circle we encounter zeros in increasing densities. The closer to the unit circle they are, the wilder and stranger they move. They move from and back to the horizon, where they are created and destroyed through bizarre bifurcations. For some special values of the parameter the unit circle suddenly gets transparent and and we get (infinitely) short glimpses of another world beyond the horizon.

We end this section by deriving eqs (E.5) and (E.6). The impenetrable prose is hopefully explained by the accompanying tables.

We know one thing from chapter **??**, 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 E.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 E.1 and the ordered list of periodic orbits (on maximal form) are intimately related. In table E.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 E.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 E.3: Example of a step in the iterative construction of the list of kneading sequences PC.

Table E.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 E.3. This fact makes it possible to generate the list of kneading sequences in an iterative way.

The zeta function at the accumulation point $H^\infty(P_1)$ is

$$\zeta_{P_1}^{-1}(z)\Xi(z^{n_1}) \quad , \tag{E.5}$$

and just before $A^\infty(P_2)$

$$\zeta_{P_2}^{-1}(z)/(1-z^{n_2}) \quad . \tag{E.6}$$

A short calculation shows that this is exactly what one would obtain by applying (E.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 (E.5) and (E.6), see Table E.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 **E.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 [16]. 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 [**?**].

E.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

$$f^n(x_i) = f(f(\dots f(x_i)\dots)) = x_i\,, \quad i = 0, 1, 2, \dots, n-1 \tag{E.7}$$

The nth 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:

$$S = (s_1 s_2 s_3 \dots s_n)^\infty = \overline{s_1 s_2 s_3 \dots s_n}\,.$$

If $\overline{s_1 s_2 \dots s_n}$ is the symbol string associated with x_0, its cyclic permutation $\overline{s_k s_{k+1} \dots s_n s_1 \dots s_{k-1}}$ corresponds to the point x_{k-1} in the same cycle. A cycle p is called *prime* if its itinerary S cannot be written as a repetition of a shorter block S'.

Each cycle yields n rational values of γ. It follows from (**??**) that if the repeating string $s_1, s_2, \dots s_n$ contains an odd number "1"s, the string of well ordered symbols $w_1 w_2 \dots w_n$ has to be of the double length before it repeats itself. The value γ is a geometrical sum which we can write as the finite sum

$$\gamma(\overline{s_1 s_2 \dots s_n}) = \frac{2^{2n}}{2^{2n}-1} \sum_{t=1}^{2n} w_t/2^t$$

Using this we can calculate the $\hat{\gamma}(S)$ for all short cycles. For orbits up to length 5 this is done in table **??**.

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

$$\{a_1, a_2, a_3, a_4, \cdots; \overline{0}\} = \{1, 10, 100, 1000, \dots; \overline{0}\}\,. \tag{E.8}$$

In this notation the itinerary $S = a_i a_j a_k a_l \cdots$ and the corresponding topological coordinate (10.18) are related by $\gamma(S) = .1^i 0^j 1^k 0^l \cdots$. For example:

$$\begin{array}{lcllcl} S & = & 111011101001000\dots & = & a_1 a_1 a_2 a_1 a_1 a_2 a_3 a_4 \dots \\ \gamma(S) & = & .101101001110000\dots & = & .1^1 0^1 1^2 0^1 1^1 0^2 1^3 0^4 \dots \end{array}$$

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 $\gamma(0 \dots 0 S) = \gamma(S)/2^i$.

A *periodic point* (or a *cycle* point) x_i belonging to a cycle of period n is a real solution of

$$f^n(x_i) = f(f(\dots f(x_i)\dots)) = x_i\,, \quad i = 0, 1, 2, \dots, n-1\,. \tag{E.9}$$

The nth 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:

$$S = (s_1 s_2 s_3 \dots s_n)^\infty = \overline{s_1 s_2 s_3 \dots s_n}\,.$$

As all itineraries are infinite, we shall adopt convention that a finite string itinerary $S = s_1 s_2 s_3 \dots s_n$ stands for infinite repetition of a finite block, and routinely omit the overline. If $\overline{s_1 s_2 \dots s_n}$ is the symbol string associated with x_0, its cyclic permutation $\overline{s_k s_{k+1} \dots s_n s_1 \dots s_{k-1}}$ corresponds to the point x_{k-1} in the same cycle. A 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 γ, but we have to distinguish *even* and *odd* cycles. The even (odd) cycles contain even (odd) number of a_i in the repeating block, with periodic points given by

$$\gamma(a_i a_j \cdots a_k a_\ell) = \begin{cases} \frac{2^n}{2^n-1}.1^i 0^j \cdots 1^k & \text{even} \\ \frac{1}{2^n+1}\left(1 + 2^n \times .1^i 0^j \cdots 1^\ell\right) & \text{odd} \end{cases}\,, \tag{E.10}$$

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:

$$\begin{array}{llllllll} \hat\gamma(1) & = & \gamma(a_1) & = & .10101\dots & = & .\overline{10} & = & 2/3 \\ \hat\gamma(10) & = & \gamma(a_2) & = & .1^2 0^2 \dots & = & .\overline{1100} & = & 4/5 \\ \hat\gamma(100) & = & \gamma(a_3) & = & .1^3 0^3 \dots & = & .\overline{111000} & = & 8/9 \\ \hat\gamma(101) & = & \gamma(a_2 a_1) & = & .1^2 0^1 \dots & = & .\overline{110} & = & 6/7 \end{array}$$

An example of a cycle where only the third symbol determines the maximal value cycle point is

$$\hat\gamma(1101110) = \gamma(a_2 a_1 a_2 a_1 a_1) = .\overline{11011010010010} = 100/129\,.$$

Maximal values of all cycles up to length 5 are given in table!?

E.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 $\mathcal{P} = \{p_1, p_2, p_3, \cdots\}$ 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 (13.2) 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 .

$$\prod_p \frac{1}{1 - t_p} = \sum_n t_n \tag{E.11}$$

A striking aspect of this formula is its resemblance to the factorization of natural numbers into primes: the relation of the cycle expansion (E.11) to the product over prime cycles is analogous to the Riemann zeta (exercise c) represented as a sum over natural numbers vs. its Euler product representation.

We now implement this factorization explicitly by decomposing recursively binary strings into ordered concatenations of prime strings. There are 2 strings of length 1, both prime: $p_1 = 0$, $p_2 = 1$. There are 4 strings of length 2: 00, 01, 11, 10. The first three are ordered concatenations of primes: $00 = p_1^2$, $01 = p_1 p_2$, $11 = p_2^2$; by ordered concatenations we mean that $p_1 p_2$ is legal, but $p_2 p_1$ is not. The remaining string is the only prime of length 2, $p_3 = 10$. Proceeding by discarding the strings which are concatenations of shorter primes $p_1^{k_1} p_2^{k_2} \cdots p_j^{k_j}$, with primes lexically ordered, we generate the standard list of primes, in agreement with table 10.1: 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 E.4.

factors	string
p_1	0
p_2	1
p_1^2	00
p_1p_2	01
p_2^2	11
p_3	10
p_1^3	000
$p_1^2p_2$	001
$p_1p_2^2$	011
p_2^3	111
p_1p_3	010
p_2p_3	110
p_4	100
p_5	101

factors	string
p_1^4	0000
$p_1^3p_2$	0001
$p_1^2p_2^2$	0011
$p_1p_2^3$	0111
p_2^4	1111
$p_1^2p_3$	0010
$p_1p_2p_3$	0110
$p_2^2p_3$	1110
p_3^2	1010
p_1p_4	0100
p_2p_4	1100
p_1p_5	0101
p_2p_5	1101
p_6	1000
p_7	1001
p_8	1011

factors	string
p_1^5	00000
$p_1^4p_2$	00001
$p_1^3p_2^2$	00011
$p_1^2p_2^3$	00111
$p_1p_2^4$	01111
p_2^5	11111
$p_1^3p_3$	00010
$p_1^2p_2p_3$	00110
$p_1p_2^2p_3$	01110
$p_2^3p_3$	11110
$p_1p_3^2$	01010
$p_2p_3^2$	11010
$p_1^2p_4$	00100
$p_1p_2p_4$	01100
$p_2^2p_4$	11100
p_3p_4	10100

factors	string
$p_1^2p_5$	00101
$p_1p_2p_5$	01101
$p_2^2p_5$	11101
p_3p_5	10101
p_1p_6	01000
p_2p_6	11000
p_1p_7	01001
p_2p_7	11001
p_1p_8	01011
p_2p_8	11011
p_9	10000
p_{10}	10001
p_{11}	10010
p_{12}	10011
p_{13}	10110
p_{14}	10111

Table E.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 $p_1 = 0$, $p_2 = 1$, $p_3 = 10$, $p_4 = 100$, ... , $p_{14} = 10111$.

E.2.1 Prime factorization for spectral determinants

Following sect. E.2, the spectral determinant cycle expansions is obtained by expanding F as a multinomial in prime cycle weights t_p

$$F = \prod_p \sum_{k=0}^{\infty} C_{p^k} t_p^k = \sum_{k_1 k_2 k_3 \cdots = 0}^{\infty} \tau_{p_1^{k_1} p_2^{k_2} p_3^{k_3} \ldots} \tag{E.12}$$

where the sum goes over all pseudocycles. In the above we have defined

$$\tau_{p_1^{k_1} p_2^{k_2} p_3^{k_3} \ldots} = \prod_{i=1}^{\infty} C_{p_i^{k_i}} t_{p_i}^{k_i} \ . \tag{E.13}$$

C on p. 166

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. E.2, the relation of the cycle expansion (E.12) to the product formula (8.9) 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 (for example, the stability eigenvalues (**??**)) 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, $t_{p_1p_2} \approx t_{p_1}t_{p_2}$.

The coefficients C_{p^k} have a simple form only in 1-d, given by the Euler formula (9.27). In higher dimensions C_{p^k} can be evaluated by expanding (8.9), $F(z) = \prod_p F_p$, where

$$F_p = 1 - \left(\sum_{r=1}^{\infty} \frac{t_p^r}{r d_{p,r}}\right) + \frac{1}{2}\left(\sum_{r=1}^{\infty} \frac{t_p^r}{r d_{p,r}}\right)^2 - \ldots .$$

Expanding and recollecting terms, and suppressing the p cycle label for the moment, we obtain

$$\begin{aligned} F_p &= \sum_{r=1}^{\infty} C_k t^k, \quad C_k = (-)^k c_k / D_k, \\ D_k &= \prod_{r=1}^{k} d_r = \prod_{a=1}^{d}\prod_{r=1}^{k}(1-u_a^r) \end{aligned} \tag{E.14}$$

where evaluation of c_k requires a certain amount of not too luminous algebra:

$$\begin{aligned} c_0 &= 1 \\ c_1 &= 1 \\ c_2 &= \frac{1}{2}\left(\frac{d_2}{d_1} - d_1\right) = \frac{1}{2}\left(\prod_{a=1}^{d}(1+u_a) - \prod_{a=1}^{d}(1-u_a)\right) \\ c_3 &= \frac{1}{3!}\left(\frac{d_2 d_3}{d_1^2} + 2d_1 d_2 - 3d_3\right) \\ &= \frac{1}{6}\left(\prod_{a=1}^{d}(1+2u_a+2u_a^2+u_a^3)\right. \\ &\qquad \left. +2\prod_{a=1}^{d}(1-u_a-u_a^2+u_a^3) - 3\prod_{a=1}^{d}(1-u_a^3)\right) \end{aligned}$$

etc.. For example, for a general 2-dimensional map we have

$$F_p = 1 - \frac{1}{D_1}t + \frac{u_1+u_2}{D_2}t^2 - \frac{u_1u_2(1+u_1)(1+u_2)+u_1^3+u_2^3}{D_3}t^3 + \ldots . \tag{E.15}$$

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. E.2.

How is the factorization of sect. E.2 used in practice? Suppose we have computed (or perhaps even measured in an experiment) all prime cycles up to length n, that is we have a list of t_p's and the corresponding Jacobian matrix eigenvalues $\Lambda_{p,1}, \Lambda_{p,2}, \ldots \Lambda_{p,d}$. 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

$$F = \sum_n \tau_n \tag{E.16}$$

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 E.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 τ_{01^3}, τ_{001^2} are known functions of the corresponding cycle eigenvalues. The zeros of F can now be easily determined by standard numerical methods. The fact that as far as the symbolic dynamics is concerned, the cycle expansion of a Selberg product is simply an average over all symbolic strings makes Selberg products rather pretty.

To be more explicit, we illustrate the above by expressing binary strings as concatenations of prime factors. We start by computing N_n, the number of terms in the expansion (E.12) of the total cycle length n. Setting $C_{p^k} t_p^k = z^{n_p k}$ in (E.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 (E.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 (E.12).

Appendix F

Counting itineraries

F.1 Counting curvatures

One consequence of the finitness 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 labelling (13.5),

$$\begin{aligned} c_4 &= -t_{0001} - t_{0011} - t_{0111} - t_0 t_{01} t_1 \\ &\quad + t_0 t_{001} + t_0 t_{011} + t_{001} t_1 + t_{011} t_1 \,. \end{aligned} \tag{F.1}$$

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.

F.2 on p. 641

Such counting rules arise from the identity

$$\prod_p (1 + t_p) = \prod_p \frac{1 - {t_p}^2}{1 - t_p} \,. \tag{F.2}$$

Substituting $t_p = z^{n_p}$ and using (11.14) 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

F.4 on p. 642

and $n > 1$, the number of terms contributing to c_n is $(N-1)N^{k-1}$ (of which half carry a minus sign).

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 cancellations between them.

Exercises

F.1 Lefschetz zeta function. Elucidate the relation betveen the topological zeta function and the Lefschetz zeta function. This should help make sense of sect. **??**.

F.2 Counting the 3-disk pinball counterterms. Verify that the number of terms in the 3-disk pinball curvature expansion (13.30) is given by

$$\begin{aligned}\prod_p (1+t_p) &= \frac{1-3z^4-2z^6}{1-3z^2-2z^3} = 1+3z^2+2z^3+\frac{z^4(6+12z+2z^2)}{1-3z^2-2z^3} \\ &= 1+3z^2+2z^3+6z^4+12z^5+20z^6+48z^7+84z^8+184z^9+\end{aligned} \quad \text{(F..3)}$$

This means that, for example, c_6 has a total of 20 terms, in agreement with the explicit 3-disk cycle expansion (13.31).

F.3 Cycle expansion denominators****.** Prove that the denominator of c_k is indeed D_k, as asserted (E.14).

F.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 (10.15) 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 (11.4) collapse to $tr(T^k) = (t_0+t_1)^k$, and $1/\zeta$ is simply

$$\prod_p (1-t_p) = 1-t_0-t_1 \tag{F.4}$$

Substituting into the identity

$$\prod_p (1+t_p) = \prod_p \frac{1-{t_p}^2}{1-t_p}$$

we obtain

$$\begin{aligned}\prod_p (1+t_p) &= \frac{1-t_0^2-t_1^2}{1-t_0-t_1} = 1+t_0+t_1+\frac{2t_0t_1}{1-t_0-t_1} \\ &= 1+t_0+t_1+\sum_{n=2}^{\infty}\sum_{k=1}^{n-1} 2\binom{n-2}{k-1}t_0^k t_1^{n-k}. \end{aligned} \tag{F.5}$$

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

$$\begin{aligned} M_{1,0} &= M_{1,1} = 1 \\ nM_{n,k} &= \sum_{m\left|\frac{n}{k}\right.} \mu(m)\binom{n/m}{k/m}, \quad n \geq 2\,, k = 1, \ldots, n-1 \end{aligned}$$

where the sum is over all m which divide both n and k.

Appendix G

Applications

> 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. **??**. For higher-dimensional flows only the Jacobian matrices are multiplicative, not individual eigenvalues, and the construction of the evolution operator for evaluation of the Lyapunov spectra requires the extension of evolution equations to the flow in the tangent space. We now develop the requisite theory.

Here we construct a multiplicative evolution operator (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 $\eta(0) \in TU_x$, 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 [21]:

$$\dot{x} = \mathbf{v}(x)\,, \quad \dot{\eta} = \mathbf{D}\mathbf{v}(x)\eta\,.$$

Here $\mathbf{D}\mathbf{v}(x)$ is the derivative matrix of the flow. We write the solution as

$$x(t) = f^t(x_0)\,, \quad \eta(t) = \mathbf{J}^t(x_0)\cdot\eta_0\,, \tag{G.1}$$

with the tangent space vector η transported by the stability matrix $\mathbf{J}^t(x_0) = \partial x(t)/\partial x_0$.

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 $\eta \in TU_x$; $\eta(x)\mathbf{v}(x) = 0$ to the $(d+1)$-dimensional dynamical evolution space $x \in U \subset \mathbb{R}^{d+1}$. In order to determine the length of the vector η we introduce a homogeneous differentiable scalar function $g(\eta) = ||\eta||$. It has the property $g(\Lambda\eta) = |\Lambda|g(\eta)$ for any Λ. An example is the projection of a vector to its dth 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

$$\Lambda^t(x_0, \mathbf{u}_0) = g(\eta(t)) = g(\mathbf{J}^t(x_0)\cdot\mathbf{u}_0) \tag{G.2}$$

is the multiplicative "stretching" factor.

Unlike the leading eigenvalue of the Jacobian the stretching factor is multiplicative along the trajectory:

$$\Lambda^{t'+t}(x_0, \mathbf{u}_0) = \Lambda^{t'}(x(t), \mathbf{u}(t))\,\Lambda^t(x_0, \mathbf{u}_0).$$

G.1 on p. 655

The $\mathbf{u}$ evolution constrained to $ET_{g,x}$, the space of unit transverse tangent vectors, is given by rescaling of (G.1):

$$\mathbf{u}' = R^t(x,\mathbf{u}) = \frac{1}{\Lambda^t(x,\mathbf{u})}\mathbf{J}^t(x)\cdot\mathbf{u}\,. \tag{G.3}$$

Eqs. (G.1), (G.2) and (G.3) enable us to define a *multiplicative* evolution operator on the extended space $U \times ET_{g,x}$

$$\mathcal{L}^t(x',\mathbf{u}';x,\mathbf{u}) = \delta\big(x' - f^t(x)\big)\,\frac{\delta\big(\mathbf{u}' - R^t(x,\mathbf{u})\big)}{|\Lambda^t(x,\mathbf{u})|^{\beta-1}}\,, \tag{G.4}$$

where β 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:

$$\begin{aligned}\mathrm{Tr}\,\mathcal{L}^t &= \sum_p T_p \sum_{r=1}^{\infty} \frac{\delta(t-rT_p)}{|\det\left(1-\mathbf{J}_p^r\right)|}\Delta_{p,r},\\ \Delta_{p,r} &= \int_g d\mathbf{u}\,\frac{\delta\big(\mathbf{u}-R^{T_p r}(x_p,\mathbf{u})\big)}{|\Lambda^{T_p r}(x_p,\mathbf{u})|^{\beta-1}}\,,\end{aligned} \tag{G.5}$$

where $\mathbf{J}_p$ 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

$$\mathbf{J}_p(x)\cdot\mathbf{e}_i(x) = \Lambda_{p,i}\,\mathbf{e}_i(x)\,, \quad i = 1,...,d. \qquad \text{(no summation on } i)$$

The $\int d\mathbf{u}$ integral in (G.5) picks up contributions from these periodic solutions. In order to compute the stability of the ith 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

$$\begin{aligned} \delta R^T(\mathbf{e}_i) &= \frac{\mathbf{J}\cdot\delta\mathbf{u}}{g(\mathbf{J}\cdot\mathbf{e}_i)} - \frac{\mathbf{J}\cdot\mathbf{e}_i}{g(\mathbf{J}\cdot\mathbf{e}_i)^2}\left(\frac{\partial g(\mathbf{e}_i)}{\partial\mathbf{u}}\cdot\mathbf{J}\cdot\delta\mathbf{u}\right) \\ &= \sum_{k\neq i}\frac{\Lambda_{p,k}}{\Lambda_{p,i}}\left(\mathbf{e}_k - \mathbf{e}_i\frac{\partial g(\mathbf{e}_i)}{\partial u_k}\right)\delta u_k \,. \end{aligned} \tag{G.6}$$

The δu_i component does not contribute to this sum since $g(\mathbf{e}_i + du_i\mathbf{e}_i) = 1 + du_i$ implies $\partial g(\mathbf{e}_i)/\partial u_i = 1$. Indeed, infinitesimal variations $\delta\mathbf{u}$ 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 (7.19) is a logarithmic derivative $\mathrm{Tr}\,\mathcal{L}(s) = -\frac{d}{ds}\log F(s)$ of the spectral determinant:

$$F(\beta,s) = \exp\left(-\sum_{p,r}\frac{e^{sT_p r}}{r\mid \det\left(1-\mathbf{J}_p^r\right)\mid}\Delta_{p,r}(\beta)\right). \tag{G.8}$$

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 15.

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 ν is called the fast dynamo rate. The goal of this section is to show that periodic orbit theory can be developed for such a highly non-trivial system as well.

We can write the solution of (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 Jacobian matrix of the transformation that moves the fluid into itself $\mathbf{x} = \mathbf{x}(\mathbf{a},t)$.

We write $\mathbf{x} = f^t(\mathbf{a})$, where f^t is the flow that maps the initial positions of the fluid particles into their positions at time t. Its inverse, $\mathbf{a} = f^{-t}(\mathbf{x})$, maps particles at time t and position $\mathbf{x}$ back to their initial positions. Then we can write (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 $Re(\nu_0) > Re(\nu_i)$, $i = 1,2,3,....$ 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

$$\mathbf{Tr}\mathcal{L}^t = \sum_p T_p \sum_{r=1}^{\infty} \frac{\mathrm{tr}\, \mathbf{J}_p^r}{\left|\det\left(\mathbf{1} - \mathbf{J}_p^{-r}\right)\right|} \delta(t - rT_p). \tag{G.20}$$

We can construct the corresponding spectral determinant as usual

$$F(s) = \exp\left[-\sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{\mathrm{tr}\, \mathbf{J}_p^r}{\left|\det\left(\mathbf{1} - \mathbf{J}_p^{-r}\right)\right|} e^{srT_p} \right] \quad . \tag{G.21}$$

Note that in this formuli we have omitted a term arising from the Jacobian transformation along the orbit which would give $1 + \mathrm{tr}\ \mathbf{J}_p^r$ in the numerator rather than just the trace of $\mathbf{J}_p^r$. 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 Λ_p (with $|\Lambda_p| > 1$), and

one contracting eigenvalue λ_p (with $|\lambda_p| < 1$) the determinant may be expanded as follows:

$$\left|\det\left(\mathbf{1}-\mathbf{J}_p^{-r}\right)\right|^{-1} = |(1-\Lambda_p^{-r})(1-\lambda_p^{-r})|^{-1} = |\lambda_p|^r \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \Lambda_p^{-jr}\lambda_p^{kr} \quad . \text{(G.22)}$$

With this decomposition we can rewrite the exponent in (G.21) as

$$\sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{(\lambda_p^r + \Lambda_p^r)e^{srT_p}}{\left|\det\left(\mathbf{1}-\mathbf{J}_p^{-r}\right)\right|} = \sum_p \sum_{j,k=0}^{\infty}\sum_{r=1}^{\infty} \frac{1}{r}\left(|\lambda_p|\Lambda_p^{-j}\lambda_p^k e^{sT_p}\right)^r (\lambda_p^r+\Lambda_p^r) \quad , \text{(G.23)}$$

which has the form of the expansion of a logarithm:

$$\sum_p \sum_{j,k} \left[\log\left(1-e^{sT_p}|\lambda_p|\Lambda_p^{1-j}\lambda_p^k\right) + \log\left(1-e^{sT_p}|\lambda_p|\Lambda_p^{-j}\lambda_p^{1+k}\right)\right] \quad . \text{(G.24)}$$

The spectral determinant is therefore of the form,

$$F(s) = F_e(s)F_c(s) \quad , \text{(G.25)}$$

where

$$F_e(s) = \prod_p \prod_{j,k=0}^{\infty}\left(1-t_p^{(jk)}\Lambda_p\right) \quad , \text{(G.26)}$$

$$F_c(s) = \prod_p \prod_{j,k=0}^{\infty}\left(1-t_p^{(jk)}\lambda_p\right), \text{(G.27)}$$

with

$$t_p^{(jk)} = e^{sT_p}|\lambda_p|\frac{\lambda_p^k}{\Lambda_p^j} \quad . \text{(G.28)}$$

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 three-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.

G.2 on p. 655

Commentary

Remark **G.1** Dynamo zeta. The dynamo zeta (G.34) has been introduced by Aurell and Gilbert [3] and reviewed in ref. [5]. Our exposition follows ref. [4].

References

[G.1] J. H. Hannay and A. M. Ozorio de Almeida, *J. Phys.* **A 17**, 3429, (1984).

[G.2] Ya.B. Pesin, *Uspekhi Mat. Nauk* **32,** 55 (1977), [*Russian Math. Surveys* **32,** 55 (1977)].

[G.3] E. Aurell and A. Gilbert, *Geophys. & Astrophys. Fluid Dynamics* (1993).

[G.4] N.J. Balmforth, P. Cvitanović, G.R. Ierley, E.A. Spiegel and G. Vattay, *Stochastic Processes in Astrophysics*, *Annals of New York Academy of Sciences* **706**, 148 (1993).

[G.5] S. Childress and A.D. Gilbert *Stretch, Twist, Fold: The Fast Dynamo*, *Lecture Notes in Physics* **37** (Springer Verlag, Berlin 1995).

[G.6] J. Balatoni and A. Renyi, *Publi. Math. Inst. Hung. Acad. Sci.* **1**, 9 (1956); (english translation **1**, 588 (Akademia Budapest, 1976)).

[G.7] R. Benzi, G. Paladin, G. Parisi and A. Vulpiani, *J. Phys.* **A17**, 3521 (1984).

[G.8] Even though the thermodynamic formalism is of older vintage (we refer the reader to ref. [16] for a comprehensive overview), we adhere here to the notational conventions of ref. [9] which are more current in the physics literature.

[G.9] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia and B.I. Shraiman, *Phys. Rev.* **A107**, 1141 (1986).

[G.10] 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.11] C. Grebogi, E. Ott and J.A. Yorke, *Phys. Rev.* **A 36**, 3522 (1987).

[G.12] C. Grebogi, E. Ott and J. Yorke, *Physica* **D 7**, 181 (1983).

[G.13] C. Grebogi, E. Ott and J.A. Yorke, *Phys. Rev.* **A36**, 3522 (1987).

[G.14] C. Grebogi, E. Ott and J. Yorke, *Phys. Rev. A* **37,** 1711 (1988).

[G.15] E. Ott, C. Grebogi and J.A. Yorke, *Phys. Lett. A* **135**, 343 (1989).

[G.16] P. Grassberger and I. Procaccia, *Phys. Rev. A* **31**, 1872 (1985).

[G.17] C. Shannon, *Bell System Technical Journal,* **27**, 379 (1948).

[G.18] H. Fujisaka, *Progr. Theor. Phys* **70**, 1264 (1983).

[G.19] A. Politi, R. Badii and P. Grassberger, *J. Phys. A* **15**, L763 (1988); P. Grassberger, R. Badii and A. Politi, *J. Stat. Phys.* **51**, 135 (1988).

[G.20] M. Barnsley, *Fractals Everywhere*, (Academic Press, New York 1988).

[G.21] J. Balatoni and A. Renyi, *Publi. Math. Inst. Hung. Acad. Sci.* **1**, 9 (1956); (english translation **1**, 588 (Akademia Budapest, 1976)).

[G.22] J.D. Crawford and J.R. Cary, *Physica* **D6**, 223 (1983)

[G.23] P. Collet and S. Isola, *Commun.Math.Phys.* **139**, 551 (1991)

[G.24] F. Christiansen, S. Isola, G. Paladin and H.H. Rugh, *J. Phys.* **A 23**, L1301 (1990).

[G.25] A.S. Pikovsky, unpublished.

[G.26] C. Beck, "Higher correlation functions of chaotic dynamical systems - a graph theoretical approach", *Nonlinearity* **4**, 1131 (1991); to be published.

[G.27] The 4-point correlation function is given in ref. [26].

[G.28] G. Hackenbroich and F. von Oppen, "Semiclassical theory of transport in antidot lattices", *Z. Phys.* **B 97**, 157 (1995).

[G.29] M.J. Feigenbaum, *J. Stat. Phys.* **21**, 669 (1979); reprinted in ref. [4].

[G.30] P. Szépfalusy, T. Tél, A. Csordás and Z. Kovács, *Phys. Rev.* **A 36**, 3525 (1987).

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?

Appendix H

Discrete symmetries

H.1 Preliminaries and Definitions

(P. Cvitanović and A. Wirzba)

In the following we will define what we mean by the concepts *group*, *representation*, *symmetry of a dynamical system*, and *invariance*.

Group axioms. First, we define a group in abstract terms: A group G is a set of elements $g_1, g_2, g_3, \ldots$ for which a law of *composition* or *group multiplication* is given such that the product $g_2 \circ g_1$ (which we will also just 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 an element e called *identity* such that $g \circ e = e \circ g = g$ for every element $g \in G$.
4. For every element $g \in G$, there exists an unique element $h \in G$ such that $h \circ g = g \circ h = e$. The element h is called *inverse* of g, and is denoted by $h = g^{-1}$.

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, will be referred to as *order* of the group.

Matrix group on vector space. We will now apply these abstract group definitions to the set of $[d \times d]$-dimensional non-singular matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \ldots$ acting in a d-dimensional vector space $V \in \mathbb{R}^d$, that is the product of matrices $\mathbf{A}$ and $\mathbf{B}$ gives the single matrix $\mathbf{C}$, such that

$$\mathbf{C}v = \mathbf{B}(\mathbf{A}v) \in V, \qquad \forall v \in V. \tag{H.1}$$

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.

Linear representation of a group. Let us now map the abstract group G *homeomorphically* on a group of matrices $\mathsf{D}(G)$ in the vector space V, that isin 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 the so defined matrix group $\mathsf{D}(G)$ a linear or matrix *representation* of the group G in the *representation space* V. Note that the matrix operation on a vector is by definition linear. We use the specification *linear* in order to discriminate the matrix representations from other operator 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* or has the *degree* d. The matrices $\mathbf{D}(g) \in \mathsf{D}(G)$ are non-singular $[d \times d]$ matrices, which we will also just abbreviate as $\mathbf{g}$, that is$x' = \mathbf{g}x$ corresponds to the normal matrix operation $x'_i = \sum_{j=1}^{d} (\mathbf{g})_{ij} x_j = \sum_{j=1}^{d} g_{ij} x_j$.

Character of a representation. The character of $\chi_\alpha(g)$ of an d-dimensional representation $\mathbf{D}(g)$ of the group element g of a discrete group G is defined as trace

$$\chi_\alpha(g) = \sum_{i=1}^{n} \mathbf{D}_{ii}(g) \equiv \mathrm{tr}\, \mathbf{D}(g)\,.$$

Note especially that $\chi(e) = n$, since $\mathbf{D}_{ij}(e) = \delta_{ij}$ for $1 \leq i, j \leq n$.

Faithful representations. If the homomorphismus mapping G on $\mathsf{D}(G)$ becomes 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 of G that will be mapped on the unit matrix $\mathbf{D}(e) = \mathbf{1}$. 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 cosidered representation is a faithful representation of the *factor group* G/H.

Equivalent representations. From this remarks it should be clear that the 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 $\mathbf{D}'(g)$. In this case, however, the new matrices $\mathbf{D}'(g)$ and the old matrices $\mathbf{D}(g)$ are related by an equivalence transformation through a non-singular matrix $\mathbf{C}$

$$\mathbf{D}'(g) = \mathbf{C}\,\mathbf{D}(g)\,\mathbf{C}^{-1}\,.$$

Thus, the group of matrices $\mathbf{D}'(g)$ form an equivalent representation $\mathsf{D}'(G)$ 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 and because equivalent representations have the same dimension, the character of equivalent representations is the same

$$\chi(g) = \sum_{i=1}^{n} \mathbf{D}'_{ii}(g) = \mathrm{tr}\,\left(\mathbf{D}'(g)\right) = \mathrm{tr}\,\left(\mathbf{C}\mathbf{D}(g)\mathbf{C}^{-1}\right)\,.$$

Regular representation of a group. The *regular* representation of a group is a special representation that is defined as follows: If we define the elements of a finite group as $g_1, g_2, \ldots, g_{|G|}$, the multiplying from the left by any element g_ν permutes the $g_1, g_2, \ldots, g_{|G|}$ among themselves. We can represent the element g_ν by the permutations of the $|G|$ "coordinates" $g_1, g_2, \ldots, g_{|G|}$. 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 this regular representation the diagonal elements of all matrices are zero except for the element g_{ν_0} with $g_{\nu_0} g_i = g_i$, that isfor the identity element e. So in the regular representation the character is given by

$$\chi(g) = \begin{cases} 1 & \text{for} \quad g = e\,, \\ 0 & \text{for} \quad g \neq e\,. \end{cases}$$

Passive and active coordinate transformations. We have to discriminate between *active* and *passive* coordinate transformations. An active (coordinate) transformation corresponds to an non-singular $d \times d$ matrix that actively shifts/changes the vector $x \in \mathcal{M}$

$$x \to \mathbf{T}x.$$

The corresponding passive coordinate transformation changes the coordinate system with respect to which the vector $f(x) \in \mathcal{M}$ is measured. Thus it is given by

$$f(x) \to \mathbf{T}^{-1} f(x) = f(\mathbf{T}^{-1} x).$$

Note that the combination of an passive and active coordinate transformation results to the identity

$$f(x) = \mathbf{T}^{-1} f(\mathbf{T}x)\,.$$

On the other hand, the evolution operator $\mathcal{L}(x, y)$ satisfies the following identity

$$\mathcal{L}(x, y) = \left| \det \left(\frac{\partial \mathbf{T}x}{\partial x} \right) \right| \mathcal{L}(\mathbf{T}x, \mathbf{T}y) = |\det \mathbf{T}|\, \mathcal{L}(\mathbf{T}x, \mathbf{T}y).$$

Note the appearance of $\det \mathbf{T}$ instead of $\det \mathbf{T}^{-1}$ and therefore the contravariant transformation property of $\mathcal{L}(x, y)$ in correspondence to maps $f(x)$. If the coordinate transformation $\mathbf{T}$ belongs to the linear non-singular representation of a *discrete* (that isfinite) symmetry group G, then $|\det \mathbf{T}| = 1$, since for any element g of a finite group G, where exists a number m such that

$$g^n \equiv \underbrace{g \circ g \circ \ldots \circ g}_{m \text{ times}} = e.$$

Thus $\mathbf{T}$ corresponds to the m-th root of $\mathbf{1}$ and the modulus of its determinant is unity.

Symmetry of dynamical system. A dynamical system $(\mathcal{M}, f)$ is invariant under a discrete symmetry group $G = \{e, g_2, \ldots, g_{|G|}\}$, if the map $f : \mathcal{M} \to \mathcal{M}$ (or the continous flow f^t) from the d-dimensional manifold $\mathcal{M}$ into itself (with d finite) is invariant

$$f(\mathbf{g}x) = \mathbf{g}f(x)$$

for *any* coordinate $x \in \mathcal{M}$ and *any* finite non-singular linear representation (that isa non-singular $d \times d$ matrix) $\mathbf{g}$ of any element $g \in G$. So a *symmetry for a dynamical system* $(\mathcal{M}, f)$ has to satisfy the two conditions

1) $\mathbf{g}x \in \mathcal{M} \quad \forall x \in \mathcal{M}$ and $\forall g \in G$,
2) $[\mathbf{D}(g), f] = 0 \quad \forall f : \mathcal{M} \to \mathcal{M}$ and $\forall g \in G$.

Group integration. Note the following laws

$$\frac{1}{|G|} \sum_{g \in G} = 1$$

and therefore

$$\frac{1}{|G|} \sum_{g \in G} \mathbf{D}(g_i) = \mathbf{D}(g_{i_0}), \quad i_0 \text{ fixed} .$$

However,

$$\frac{1}{|G|} \sum_{g \in G} \mathbf{D}(g) = \mathbf{0},$$

where $\mathbf{0}$ is the zero matrix of same dimension as the representations $\mathbf{D}(g) \in D(G)$. In particular,

$$\frac{1}{|G|} \sum_{g \in G} \chi_\alpha(g) = \frac{1}{|G|} \sum_{g \in G} \sum_{i=1}^{d_\alpha} D(g)_{ii} = 0.$$

Furthermore, if we consider all non-equilavent irreducible representations of a group G, then the quantities $D^{(\alpha)}_{ij}(g)$ for fixed α, i and j

Orthonormalitity of characters. But what can we say about

$$\frac{1}{|G|}\sum_{g\in G}\chi_\alpha(hg)\chi_\alpha(g^{-1}k^{-1}) \quad \text{with } h,k\in G \text{ fixed ?}$$

Note the following relation

$$\delta_{ab}\delta_{cd}=\frac{1}{n}\delta_{ad}\delta_{cb}+\left(\delta_{ab}\delta_{cd}-\frac{1}{n}\delta_{ad}\delta_{cb}\right).$$

Projection operators. The projection operator onto the α irreducible subspace of dimension d_α is given by

$$\mathbf{P}_\alpha=\frac{d_\alpha}{|G|}\sum_{g\in G}\chi_\alpha(g)\mathbf{g}^{-1}.$$

Note that $\mathbf{P}_\alpha$ is a $[d\times d]$-dimensional matrix as the representation $\mathbf{g}$.

Irreducible subspaces of the evolution operator.

$$\mathcal{L}=\sum_\alpha \mathrm{tr}\,\mathcal{L}_\alpha$$

with

$$\mathcal{L}_\alpha(y,x)=\frac{d_\alpha}{|G|}\sum_{g\in G}\chi_\alpha(g)\mathcal{L}(\mathbf{g}^{-1}y,x),$$

where the prefactor d_α reflects the fact that a d_α-dimensional representation occurs d_α times.

H.2 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 $\rho_i=\epsilon_{i+1}-\epsilon_i \mod N$ 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 (fig. H.1a). The symmetry group consists of the identity **e**, the two reflections σ_x, σ_y across x, y axes, the two diagonal reflections σ_{13}, σ_{24}, and the three rotations C_4, C_2 and C_4^3 by angles $\pi/2$, π 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 $\to$ 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	$\leftrightarrow$	$\underline{1}$
1234	$\leftrightarrow$	1
13	$\leftrightarrow$	2

Conversions for all periodic orbits of reduced symbol period less than 5 are listed in table H.1.

This symbolic dynamics is closely related to the group-theoretic structure of the dynamics: the global 4-disk trajectory can be generated by mapping the fundamental domain trajectories onto the full 4-disk space by the accumulated product of the 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 H.1. Our convention is to multiply group elements in the reverse order with respect to the symbol sequence. We need these group elements for our next step, the dynamical zeta function factorizations.

The C_{4v} group has four one-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 two-dimensional representations E. Substituting the C_{4v} characters

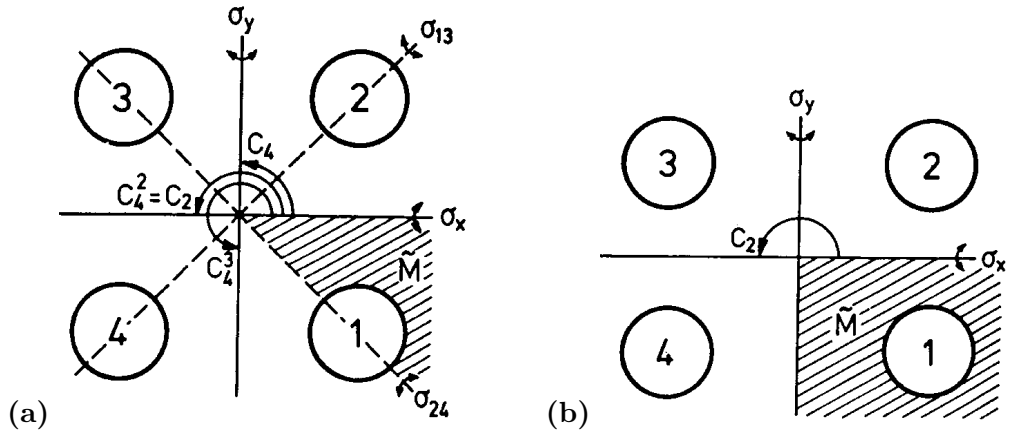


Figure H.1: (a) The symmetries of four disks on a square. (b) The symmetries of four disks on a rectangle. The fundamental domains are indicated by the shaded wedges.

$\tilde{p}$	p	$\mathbf{h}_{\tilde{p}}$
0	1 2	σ_x
1	1 2 3 4	C_4
2	1 3	C_2, σ_{13}
01	12 14	σ_{24}
02	12 43	σ_y
12	12 41 34 23	C_4^3
001	121 232 343 414	C_4
002	121 343	C_2
011	121 434	σ_y
012	121 323	σ_{13}
021	124 324	σ_{13}
022	124 213	σ_x
112	123	e
122	124 231 342 413	C_4

$\tilde{p}$	p	$\mathbf{h}_{\tilde{p}}$
0001	1212 1414	σ_{24}
0002	1212 4343	σ_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	σ_{13}
0112 (b)	1214 2123	σ_x
0121 (b)	1213 2124	σ_x
0122	1213 1413	σ_{24}
0211	1243 2134	σ_x
0212	1243 1423	σ_{24}
0221	1242 1424	σ_{24}
0222	1242 4313	σ_y
1112	1234 2341 3412 4123	C_4
1122	1231 3413	C_2
1222	1242 4131 3424 2313	C_4^3

Table H.1: C_{4v} correspondence between the ternary fundamental domain prime cycles $\tilde{p}$ and the full 4-disk {1,2,3,4} labelled 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.2 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 π and a reflection across a diagonal. The two pairs of cycles marked by (a) and (b) are related by time reversal, but cannot be mapped into each other by C_{4v} transformations.

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
σ_{axes}	1	-1	1	-1	0
σ_{diag}	1	-1	-1	1	0

into (17.15) we obtain:

$$\begin{array}{rrcccccc}
h_{\tilde{p}} & & & A_1 & A_2 & B_1 & B_2 & E \\
e\text{:} & (1-t_{\tilde{p}})^8 & = & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}})^4 \\
C_2\text{:} & (1-t_{\tilde{p}}^2)^4 & = & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}})^4 \\
C_4, C_4^3\text{:} & (1-t_{\tilde{p}}^4)^2 & = & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1+t_{\tilde{p}}^2)^2 \\
\sigma_{axes}\text{:} & (1-t_{\tilde{p}}^2)^4 & = & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1-t_{\tilde{p}}^2)^2 \\
\sigma_{diag}\text{:} & (1-t_{\tilde{p}}^2)^4 & = & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}^2)^2
\end{array}$$

The possible irreducible segment group elements $\mathbf{h}_{\tilde{p}}$ are listed in the first column; σ_{axes} denotes a reflection across either the x-axis or the y-axis, and σ_{diag} denotes a reflection across a diagonal (see fig. H.1a). 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:

$$\begin{array}{rrcl}
 & & & A_1 \quad A_2 \quad B_1 \quad B_2 \quad E \\
\text{axes:} & (1-t_{\tilde{p}}^2)^2 & = & (1-t_{\tilde{p}})(1-0t_{\tilde{p}})(1-t_{\tilde{p}})(1-0t_{\tilde{p}})(1+t_{\tilde{p}})^2 \\
\text{diagonals:} & (1-t_{\tilde{p}}^2)^2 & = & (1-t_{\tilde{p}})(1-0t_{\tilde{p}})(1-0t_{\tilde{p}})(1-t_{\tilde{p}})(1+t_{\tilde{p}})^2
\end{array} \tag{H.2}$$

(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

$$\begin{aligned}
1/\zeta_{A_1} &= (1-t_0)(1-t_1)(1-t_2)(1-t_{01})(1-t_{02})(1-t_{12}) \\
&\quad (1-t_{001})(1-t_{002})(1-t_{011})(1-t_{012})(1-t_{021})(1-t_{022})(1-t_{112}) \\
&\quad (1-t_{122})(1-t_{0001})(1-t_{0002})(1-t_{0011})(1-t_{0012})(1-t_{0021})\dots \\
&= 1-t_0-t_1-t_2-(t_{01}-t_0t_1)-(t_{02}-t_0t_2)-(t_{12}-t_1t_2) \\
&\quad -(t_{001}-t_0t_{01})-(t_{002}-t_0t_{02})-(t_{011}-t_1t_{01}) \\
&\quad -(t_{022}-t_2t_{02})-(t_{112}-t_1t_{12})-(t_{122}-t_2t_{12}) \\
&\quad -(t_{012}+t_{021}+t_0t_1t_2-t_0t_{12}-t_1t_{02}-t_2t_{01})\dots
\end{aligned} \tag{H.3}$$

(for typographical convenience, $\underline{1}$ is replaced by 0 in the remainder of this section). For one-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:

$$\begin{aligned}
1/\zeta_{B_2} &= (1+t_0)(1+t_1)(1-t_2)(1-t_{01})(1+t_{02})(1+t_{12}) \\
&\quad (1+t_{001})(1-t_{002})(1+t_{011})(1-t_{012})(1-t_{021})(1+t_{022})(1-t_{112}) \\
&\quad (1+t_{122})(1-t_{0001})(1+t_{0002})(1-t_{0011})(1+t_{0012})(1+t_{0021})\dots \\
&= 1+t_0+t_1-t_2-(t_{01}-t_0t_1)+(t_{02}-t_0t_2)+(t_{12}-t_1t_2) \\
&\quad +(t_{001}-t_0t_{01})-(t_{002}-t_0t_{02})+(t_{011}-t_1t_{01}) \\
&\quad +(t_{022}-t_2t_{02})-(t_{112}-t_1t_{12})+(t_{122}-t_2t_{12}) \\
&\quad -(t_{012}+t_{021}+t_0t_1t_2-t_0t_{12}-t_1t_{02}-t_2t_{01})\dots
\end{aligned} \tag{H.4}$$

The form of the remaining cycle expansions depends crucially on the special role played by the boundary orbits: by (H.2) the orbit t_2 does not contribute to A_2 and B_1,

$$\begin{aligned}
1/\zeta_{A_2} &= (1+t_0)(1-t_1)(1+t_{01})(1+t_{02})(1-t_{12}) \\
&\quad (1-t_{001})(1-t_{002})(1+t_{011})(1+t_{012})(1+t_{021})(1+t_{022})(1-t_{112}) \\
&\quad (1-t_{122})(1+t_{0001})(1+t_{0002})(1-t_{0011})(1-t_{0012})(1-t_{0021})\dots \\
&= 1+t_0-t_1+(t_{01}-t_0t_1)+t_{02}-t_{12} \\
&\quad -(t_{001}-t_0t_{01})-(t_{002}-t_0t_{02})+(t_{011}-t_1t_{01}) \\
&\quad +t_{022}-t_{122}-(t_{112}-t_1t_{12})+(t_{012}+t_{021}-t_0t_{12}-t_1t_{02})\dots
\end{aligned} \tag{H.5}$$

and

$$\begin{aligned}
1/\zeta_{B_1} &= (1-t_0)(1+t_1)(1+t_{01})(1-t_{02})(1+t_{12}) \\
&\quad (1+t_{001})(1-t_{002})(1-t_{011})(1+t_{012})(1+t_{021})(1-t_{022})(1-t_{112}) \\
&\quad (1+t_{122})(1+t_{0001})(1-t_{0002})(1-t_{0011})(1+t_{0012})(1+t_{0021})\dots \\
&= 1-t_0+t_1+(t_{01}-t_0t_1)-t_{02}+t_{12} \\
&\quad +(t_{001}-t_0t_{01})-(t_{002}-t_0t_{02})-(t_{011}-t_1t_{01}) \\
&\quad -t_{022}+t_{122}-(t_{112}-t_1t_{12})+(t_{012}+t_{021}-t_0t_{12}-t_1t_{02})\dots
\end{aligned} \tag{H.6}$$

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

$$1/\zeta_E = 1+t_2+(-{t_0}^2+{t_1}^2)+(2t_{002}-t_2{t_0}^2-2t_{112}+t_2{t_1}^2)$$

$$
\begin{aligned}
&+(2t_{0011} - 2t_{0022} + 2t_2t_{002} - {t_{01}}^2 - {t_{02}}^2 + 2t_{1122} - 2t_2t_{112} \\
&+{t_{12}}^2 - {t_0}^2{t_1}^2) + (2t_{00002} - 2t_{00112} + 2t_2t_{0011} - 2t_{00121} - 2t_{00211} \\
&+2t_{00222} - 2t_2t_{0022} + 2t_{01012} + 2t_{01021} - 2t_{01102} - t_2{t_{01}}^2 + 2t_{02022} \\
&-t_2{t_{02}}^2 + 2t_{11112} - 2t_{11222} + 2t_2t_{1122} - 2t_{12122} + t_2{t_{12}}^2 - t_2{t_0}^2{t_1}^2 \\
&+2t_{002}(-{t_0}^2 + {t_1}^2) - 2t_{112}(-{t_0}^2 + {t_1}^2))
\end{aligned}
\tag{H.7}
$$

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 (11.41).

13.9 on p. 317

Remark **H.1** Labelling conventions While there is a variety of labelling conventions [13, 13] 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.

H.3 C_{2v} factorization

An arrangement of four identical disks on the vertices of a rectangle has C_{2v} symmetry (fig. H.1b). C_{2v} consists of $\{e, \sigma_x, \sigma_y, C_2\}$, *i.e.*, the reflections across the symmetry axes and a rotation by π.

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 $\ldots g_1g_0g_2g_0g_1g_0g_0$.

Short ternary cycles and the corresponding 4-disk cycles are listed in table H.2. Note that already at length three there is a pair of cycles (012 = 143 and 021 = 142) related by time reversal, but *not* by any C_{2v} symmetries.

$\tilde{p}$	p	$\mathbf{g}$
0	1 4	σ_y
1	1 2	σ_x
2	1 3	C_2
01	14 32	C_2
02	14 23	σ_x
12	12 43	σ_y
001	141 232	σ_x
002	141 323	C_2
011	143 412	σ_y
012	143	e
021	142	e
022	142 413	σ_y
112	121 343	C_2
122	124 213	σ_x

$\tilde{p}$	p	$\mathbf{g}$
0001	1414 3232	C_2
0002	1414 2323	σ_x
0011	1412	e
0012	1412 4143	σ_y
0021	1413 4142	σ_y
0022	1413	e
0102	1432 4123	σ_y
0111	1434 3212	C_2
0112	1434 2343	σ_x
0121	1431 2342	σ_x
0122	1431 3213	C_2
0211	1421 2312	σ_x
0212	1421 3243	C_2
0221	1424 3242	C_2
0222	1424 2313	σ_x
1112	1212 4343	σ_y
1122	1213	e
1222	1242 4313	σ_y

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.3, starting from disk 1.

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 one-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
σ_x	1	-1	1	-1
σ_y	1	-1	-1	1

Substituted into the factorized determinant (17.14), the contributions of pe-

riodic orbits split as follows

$g_{\tilde{p}}$			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:	$(1-t_{\tilde{p}}^2)^2$	=	$(1-t_{\tilde{p}})$	$(1-t_{\tilde{p}})$	$(1-t_{\tilde{p}})$	$(1-t_{\tilde{p}})$
σ_x:	$(1-t_{\tilde{p}}^2)^2$	=	$(1-t_{\tilde{p}})$	$(1+t_{\tilde{p}})$	$(1-t_{\tilde{p}})$	$(1+t_{\tilde{p}})$
σ_y:	$(1-t_{\tilde{p}}^2)^2$	=	$(1-t_{\tilde{p}})$	$(1+t_{\tilde{p}})$	$(1+t_{\tilde{p}})$	$(1-t_{\tilde{p}})$

Cycle expansions follow by substituting cycles and their group theory factors from table H.2. For A_1 all characters are $+1$, and the corresponding cycle expansion is given in (H.3). Similarly, the totally antisymmetric subspace factorization A_2 is given by (H.4), 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:

$$\begin{aligned}
1/\zeta_{B_1} &= (1+t_0)(1-t_1)(1+t_2)(1+t_{01})(1-t_{02})(1+t_{12}) \\
&\quad (1-t_{001})(1+t_{002})(1+t_{011})(1-t_{012})(1-t_{021})(1+t_{022})(1+t_{112}) \\
&\quad (1-t_{122})(1+t_{0001})(1-t_{0002})(1-t_{0011})(1+t_{0012})(1+t_{0021})\dots \\
&= 1+t_0-t_1+t_2+(t_{01}-t_0t_1)-(t_{02}-t_0t_2)+(t_{12}-t_1t_2) \\
&\quad -(t_{001}-t_0t_{01})+(t_{002}-t_0t_{02})+(t_{011}-t_1t_{01}) \\
&\quad +(t_{022}-t_2t_{02})+(t_{112}-t_1t_{12})-(t_{122}-t_2t_{12}) \\
&\quad -(t_{012}+t_{021}+t_0t_1t_2-t_0t_{12}-t_1t_{02}-t_2t_{01})\dots
\end{aligned} \tag{H.8}$$

For B_2 all t_p with an odd total number of 1's and 2's change sign:

$$\begin{aligned}
1/\zeta_{B_2} &= (1-t_0)(1+t_1)(1+t_2)(1+t_{01})(1+t_{02})(1-t_{12}) \\
&\quad (1+t_{001})(1+t_{002})(1-t_{011})(1-t_{012})(1-t_{021})(1-t_{022})(1+t_{112}) \\
&\quad (1+t_{122})(1+t_{0001})(1+t_{0002})(1-t_{0011})(1-t_{0012})(1-t_{0021})\dots \\
&= 1-t_0+t_1+t_2+(t_{01}-t_0t_1)+(t_{02}-t_0t_2)-(t_{12}-t_1t_2) \\
&\quad +(t_{001}-t_0t_{01})+(t_{002}-t_0t_{02})-(t_{011}-t_1t_{01}) \\
&\quad -(t_{022}-t_2t_{02})+(t_{112}-t_1t_{12})+(t_{122}-t_2t_{12}) \\
&\quad -(t_{012}+t_{021}+t_0t_1t_2-t_0t_{12}-t_1t_{02}-t_2t_{01})\dots
\end{aligned} \tag{H.9}$$

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 (11.41).

Remark **H.2** *C_{2v} symmetry* C_{2v} is the symmetry of several systems studied in the literature, such as the stadium billiard [**?**], and the 2-dimensional anisotropic Kepler potential [24].

H.4 Symmetries of the symbol square

• advanced section •

Depending on the type of dynamical system, the symbol square might have a variety of symmetries. Under the time reversal

$$\cdots s_{-2}s_{-1}s_0.s_1s_2s_3\cdots \longrightarrow \cdots s_3s_2s_1.s_0s_{-1}s_{-2}\cdots$$

the points in the symbol square for an orientation preserving map are symmetric across the diagonal $\gamma = \delta$, and for the orientation reversing case they are symmetric with respect to the $\gamma = 1 - \delta$ diagonal. Consequently the periodic orbits appear either in dual pairs $p = s_1s_2s_3\ldots s_n$, $\overline{p} = s_ns_{n-1}s_{n-2}\ldots s_1$, or are self-dual under time reversal, $S_p = S_{\overline{p}}$. For the orientation preserving case a self-dual cycle of odd period has at least one point on the symmetry diagonal. In particular, all fixed points lie on the symmetry diagonal. Determination of such symmetry lines can be of considerable practical utility, as it reduces some of the periodic orbit searches to 1-dimensional searches.

Remark **H.3** *Symmetries of the symbol square.* For a more detailed discussion of the symbolic dynamics symmetries, see refs. [4, **?**].

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 13.2, 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 (13.2) is not absolutely convergent in the sense of Dirichlet series of sect. 13.5, 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, fig. 1.8.

I.1.1 Tessalation of a smooth flow by cycles

One of the early high accuracy computations of π 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 nth 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 11 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 (11.12). 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, that is 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 (13.5) 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 kth return. If the two trajectories coincide in the first m symbols, at the mth 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 $x_{a\cdots} = x_{\overline{a}} + \delta x_{a\cdots}$

$$f^{T_a}(x_{\overline{a}}) - f^{T_{a\cdots}}(x_{a\cdots}) \approx \frac{\partial f^{T_a}(x_{\overline{a}})}{\partial x} \cdot \delta x_{a\cdots} = \mathbf{J}_a \cdot \delta x_{a\cdots} ,$$

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 (6.1) 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 $\overline{a}$,

$$T_a + T_b - T_{ab} \approx \mathrm{Const} \times \Lambda_a^{-j}\,, \quad |\Lambda_a \Lambda_b/\Lambda_{ab}| \approx \exp(-\mathrm{Const}/\Lambda_{ab})$$

n	$t_{ab} - t_a t_b$	$T_{ab} - (T_a + T_b)$	$\log \left[\frac{\Lambda_a \Lambda_b}{\Lambda_{ab}}\right]$	$ab - a \cdot b$
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 12.3. The ratio $\Lambda_a \Lambda_b / \Lambda_{ab}$ is approaching unity exponentially fast.

Expanding the exponentials one thus finds that this term in the cycle expansion is of the order of

$$t_{a^j b} - t_a t_{a^{j-1} b} \approx \text{Const} \times t_{a^j b} \Lambda_a^{-j} \,. \tag{I.1}$$

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.

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. 13.1.3.

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 (13.5), the series is truncated as in sect. 13.4, 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 (14.18).

Numerical results (see for example the plots of the accuracy of the cycle expansion truncations for the Hénon map in ref. [2]) 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 (14.11) 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 9 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 (6.4) 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 β derivative of both sides of (**??**)

$$\rho_\beta(y)e^{ts(\beta)} = \int_{\mathcal{M}} dx\, \delta(y - f^t(x))e^{\beta\cdot A^t(x)}\rho_\beta(x)\,,$$

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

$$\frac{\partial s}{\partial\beta}\Big|_{\beta=0} = \int_{\mathcal{M}} dy\,\rho_0(x)\,\langle a(x)\rangle\ . \tag{I.2}$$

This is the expectation value (6.12) only if the time average (6.4) equals the space average (6.9), $\langle a(x)\rangle = \langle a\rangle$, 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 (6.18) is a Cantor set, the closure of the union of all periodic orbits. Alternatively, $\mathcal{M}$ can be considered continuous, but then the measure ρ_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

4.1 on p. 94

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"(to be discussed in sect. **??**), 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.

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 nth cumulant

An immediate consequence of the exponential spacing of the eigenvalues is that the convergence of the Selberg product expansion (E.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 Jacobian matrix eigenvalues in (**??**) 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 (8.9) 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 (9.27), and the coefficients C_k in (E.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 (13.8) 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 (8.23) In this case the expansion corresponding to (9.27) is given by (9.28) 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 for trace-class and Hilbert-Schmidt matrices, the determinants over infinite dimensional matrices and possible 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

$$\mathbf{A}'(x) = \frac{d\mathbf{A}(x)}{dx}\,, \qquad A'_{ij}(x) = \frac{d}{dx}A_{ij}(x)\,. \tag{J.1}$$

Derivatives of products of matrices are evaluated by the chain rule

$$\frac{d}{dx}(\mathbf{AB}) = \frac{d\mathbf{A}}{dx}\mathbf{B} + \mathbf{A}\frac{d\mathbf{B}}{dx}\,. \tag{J.2}$$

A matrix and its derivative matrix in general do not commute

$$\frac{d}{dx}\mathbf{A}^2 = \frac{d\mathbf{A}}{dx}\mathbf{A} + \mathbf{A}\frac{d\mathbf{A}}{dx}\,. \tag{J.3}$$

The derivative of the inverse of a matrix, follows from $\frac{d}{dx}(\mathbf{A}\mathbf{A}^{-1}) = 0$:

$$\frac{d}{dx}\mathbf{A}^{-1} = \frac{1}{\mathbf{A}}\frac{d\mathbf{A}}{dx}\frac{1}{\mathbf{A}}\,. \tag{J.4}$$

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:

$$e^{\mathbf{A}} = \sum_{k=0}^{\infty}\frac{1}{k!}\mathbf{A}^k\,, \qquad \mathbf{A}^0 = \mathbf{1} \tag{J.5}$$

$$= \lim_{N\to\infty}\left(\mathbf{1} + \frac{1}{N}\mathbf{A}\right)^N \tag{J.6}$$

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

$$\det\left(e^{\mathbf{A}}\right) = \lim_{N\to\infty}\left(\det\,\left(\mathbf{1} + \mathbf{A}/N\right)\right)^N\,.$$

To the leading order in $1/N$

$$\det\,\left(\mathbf{1} + \mathbf{A}/N\right) = 1 + \frac{1}{N}\mathrm{tr}\,\mathbf{A} + O(N^{-2})\,.$$

hence

$$\det e^{\mathbf{A}} = \lim_{N\to\infty}\left(1 + \frac{1}{N}\mathrm{tr}\,\mathbf{A} + O(N^{-2})\right)^N = e^{\mathrm{tr}\,\mathbf{A}} \tag{J.7}$$

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 derivaties

$$\frac{d}{dt}\left(e^{t\mathbf{A}}\mathbf{B}e^{-t\mathbf{A}}\right) = e^{t\mathbf{A}}[\mathbf{A},\mathbf{B}]e^{-t\mathbf{A}}\,.$$

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 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.4 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, $\{\phi_n\}$:

$$\sum_n |\langle\phi_n, \mathbf{A}\phi_n\rangle| < \infty. \tag{J.10}$$

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, $\{\phi_n\}$:

$$\sum_n \|\mathbf{A}\phi_n\|^2 < \infty. \tag{J.11}$$

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.3 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

$$\phi_1 \wedge \cdots \wedge \phi_n = \frac{1}{\sqrt{n!}} \sum_{\pi \in \mathcal{P}_n} \epsilon(\pi) [\phi_{\pi(1)} \otimes \cdots \otimes \phi_{\pi(n)}] \tag{J.12}$$

where $\mathcal{P}_n$ is the group of all permutations of n letters and $\epsilon(\pi) = \pm 1$ depending on whether π is an even or odd permutation, respectively. The inner product in $\bigwedge^n(\mathcal{H})$ is given by

$$(\phi_1 \wedge \cdots \wedge \phi_n, \eta_1 \wedge \cdots \wedge \eta_n) = \det\{(\phi_i, \eta_j)\} \tag{J.13}$$

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

$$\bigwedge^n(\mathbf{A}) (\phi_1 \wedge \cdots \wedge \phi_n) = \mathbf{A}\phi_1 \wedge \cdots \wedge \mathbf{A}\phi_n \ . \tag{J.14}$$

When $n = 0$, $\bigwedge^n(\mathcal{H})$ is defined to be C and $\bigwedge^n(\mathbf{A})$ as 1: $C \to C$.

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

$$\mathrm{Tr}\left(\bigwedge^k(\mathbf{A})\right) = \sum_{i_1<\cdots<i_k} ((\phi_{i_1} \wedge \cdots \wedge \phi_{i_k}), (\mathbf{A}\phi_{i_1} \wedge \cdots \wedge \mathbf{A}\phi_{i_k})) < \infty \quad \text{(J.15)}$$

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 $\mathbf{det(1+A)}$ is defined as

$$\det(\mathbf{1} + \mathbf{A}) = \sum_{k=0}^{\infty} \mathrm{Tr}\left(\bigwedge^k(\mathbf{A})\right) \quad \text{(J.16)}$$

Properties:

Let $\mathbf{A}$ be a linear operator on a separable Hilbert space $\mathcal{H}$ and $\{\phi_j\}_1^\infty$ an orthonormal basis.

(a) $\sum_{k=0}^{\infty} \mathrm{Tr}\left(\bigwedge^k(\mathbf{A})\right)$ converges for each $\mathbf{A} \in \mathcal{J}_1$.

(b) $|\det(\mathbf{1} + \mathbf{A})| \leq \prod_{j=1}^{\infty}(1 + \mu_j(\mathbf{A}))$ where $\mu_j(\mathbf{A})$ are the *singular* values of $\mathbf{A}$, i.e., the eigenvalues of $|\mathbf{A}| = \sqrt{\mathbf{A}^\dagger\mathbf{A}}$.

(c) $|\det(\mathbf{1} + \mathbf{A})| \leq \exp(\|\mathbf{A}\|_1)$.

(d) For any $\mathbf{A}_1, \ldots, \mathbf{A}_n \in \mathcal{J}_1$, $\langle z_1, \ldots, z_n\rangle \mapsto \det\left(\mathbf{1} + \sum_{i=1}^n z_i\mathbf{A}_i\right)$ is an entire analytic function.

(e) If $\mathbf{A}, \mathbf{B} \in \mathcal{J}_1$, then

$$\begin{aligned} \det(\mathbf{1} + \mathbf{A})\det(\mathbf{1} + \mathbf{B}) &= \det(\mathbf{1} + \mathbf{A} + \mathbf{B} + \mathbf{AB}) \\ &= \det((\mathbf{1} + \mathbf{A})(\mathbf{1} + \mathbf{B})) \\ &= \det((\mathbf{1} + \mathbf{B})(\mathbf{1} + \mathbf{A}))\ . \end{aligned} \quad \text{(J.17)}$$

If $\mathbf{A} \in \mathcal{J}_1$ and $\mathbf{U}$ unitary, then

$$\det\left(\mathbf{U}^{-1}(\mathbf{1} + \mathbf{A})\mathbf{U}\right) = \det\left(\mathbf{1} + \mathbf{U}^{-1}\mathbf{A}\mathbf{U}\right) = \det(\mathbf{1} + \mathbf{A})\ . \quad \text{(J.18)}$$

(f) If $\mathbf{A} \in \mathcal{J}_1$, then $(\mathbf{1} + \mathbf{A})$ is invertible if and only if $\det(\mathbf{1} + \mathbf{A}) \neq 0$.

(g) If $\lambda \neq 0$ is an n-times degenerate eigenvalue of $\mathbf{A} \in \mathcal{J}_1$, then $\det(\mathbf{1} + z\mathbf{A})$ has a zero of order n at $z = -1/\lambda$.

(h) For any ϵ, there is a $C_\epsilon(\mathbf{A})$, depending on $\mathbf{A} \in \mathcal{J}_1$, so that $|\det(\mathbf{1} + z\mathbf{A})| \leq C_\epsilon(\mathbf{A}) \exp(\epsilon|z|)$.

(i) For any $\mathbf{A} \in \mathcal{J}_1$,

$$\det(\mathbf{1} + \mathbf{A}) = \prod_{j=1}^{N(\mathbf{A})} (1 + \lambda_j(\mathbf{A})) \tag{J.19}$$

where here and in the following $\{\lambda_j(\mathbf{A})\}_{j=1}^{N(\mathbf{A})}$ are the eigenvalues of $\mathbf{A}$ counted with algebraic multiplicity .

(j) *Lidskii's theorem:* For any $\mathbf{A} \in \mathcal{J}_1$,

$$\mathrm{Tr}(\mathbf{A}) = \sum_{j=1}^{N(\mathbf{A})} \lambda_j(\mathbf{A}) < \infty \ . \tag{J.20}$$

(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

$$\det(1 + z\mathbf{A}) = \sum_{k=0}^{\infty} z^k \sum_{1 \leq j_1 < \cdots < j_k \leq N(\mathbf{A})} \lambda_{j_1}(\mathbf{A}) \cdots \lambda_{j_k}(\mathbf{A}) < \infty. \tag{J.21}$$

(m) If $\mathbf{A} \in \mathcal{J}_1$, then for $|z|$ small (that is $|z| \max|\lambda_j(\mathbf{A})| < 1$) the series $\sum_{k=1}^{\infty} z^k \mathrm{Tr}\left((-\mathbf{A})^k\right)/k$ converges and

$$\begin{aligned} \det(1 + z\mathbf{A}) &= \exp\left(-\sum_{k=1}^{\infty} \frac{z^k}{k} \mathrm{Tr}\left((-\mathbf{A})^k\right)\right) \\ &= \exp\left(\mathrm{Tr}\ln(\mathbf{1} + z\mathbf{A})\right) \ . \end{aligned} \tag{J.22}$$

(n) *The Plemelj-Smithies formula:* Define $\alpha_m(\mathbf{A})$ for $\mathbf{A} \in \mathcal{J}_1$ by

$$\det(\mathbf{1} + z\mathbf{A}) = \sum_{m=0}^{\infty} z^m \frac{\alpha_m(\mathbf{A})}{m!} \ . \tag{J.23}$$

Then $\alpha_m(\mathbf{A})$ is given by the $m \times m$ determinant:

$$\alpha_m(\mathbf{A}) = \begin{vmatrix} \mathrm{Tr}(\mathbf{A}) & m-1 & 0 & \cdots & 0 \\ \mathrm{Tr}(\mathbf{A}^2) & \mathrm{Tr}(\mathbf{A}) & m-2 & \cdots & 0 \\ \mathrm{Tr}(\mathbf{A}^3) & \mathrm{Tr}(\mathbf{A}^2) & \mathrm{Tr}(\mathbf{A}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & 1 \\ \mathrm{Tr}(\mathbf{A}^m) & \mathrm{Tr}(\mathbf{A}^{(m-1)}) & \mathrm{Tr}(\mathbf{A}^{(m-2)}) & \cdots & \mathrm{Tr}(\mathbf{A}) \end{vmatrix} \tag{J.24}$$

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

$$\begin{aligned} c_m(\mathbf{A}) &= \frac{1}{m}\sum_{k=1}^{m}(-1)^{k+1} c_{m-k}(\mathbf{A})\,\mathrm{Tr}(\mathbf{A}^k) \quad \text{for } m \geq 1 \\ c_0(\mathbf{A}) &\equiv 1\ . \end{aligned} \tag{J.25}$$

Note that formula (J.23) 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, that is let $c_m(\text{s.c.})$ denote the $m^{\,\text{th}}$ curvature term, then the curvature expansion of the semiclassical zeta function is given by the recursion relation

$$\begin{aligned} c_m(\text{s.c.}) &= \frac{1}{m}\sum_{k=1}^{m}(-1)^{k+m+1} c_{m-k}(\text{s.c.}) \sum_{\substack{p;r>0 \\ \text{with } [p]r=k}} [p]\,\frac{t_p(k)^r}{1-\left(\frac{1}{\Lambda_p}\right)^r} \quad \text{for } m \geq 1 \\ c_0(\text{s.c.}) &\equiv 1\ . \end{aligned} \tag{J.26}$$

In fact, in the cumulant expansion (J.23) as well as in the curvature expansion there are large cancellations involved. Let us order – without lost of generality – the eigenvalues of the operator $\mathbf{A} \in \mathcal{J}_1$ as follows:

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_{i-1}| \geq |\lambda_i| \geq |\lambda_{i+1}| \geq \cdots$$

(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.23), we have enormous cancellations of big numbers, e.g. at the $k^{\,\text{th}}$ cumulant order ($k > 3$), all the intrinsically large 'numbers' λ_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 $\sum_{1\leq j_1<\cdots<j_k\leq N(\mathbf{A})} \lambda_{j_1}\cdots\lambda_{j_k}$ is finally left over. Algebraically, the fact that there are these large cancellations is of course of no importance. However, if the determinant is calculated numerically, the big cancellations might spoil the result or even the convergence. Now, the curvature expansion of the semiclassical zeta function, as it is known today, *is* the

semi-classical approximation to the curvature expansion (unfortunately) in the Plemelj-Smithies form. As the exact quantum mechanical result is approximated semi-classically, 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 semi-classical analog to the reduced cumulant expansion (J.21) or even to (J.19) directly. It might not be possible to find a direct semi-classical analog for the individual eigenvalues λ_j. Thus the direct construction of the semi-classical equivalent to (J.19) is rather unlikely. However, in order to have a semi-classical "cumulant" summation without large cancellations – see (J.21) – it would be just sufficient to find the semi-classical analog of each complete cumulant (J.21) and not of the single eigenvalues. Whether this will eventually be possible is still an open question.

J.4 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

$$\sum_{j=1}^{\infty} |a_{jj}| \quad < \quad \infty\,, \tag{J.27}$$

$$\sum_{j,k=1}^{\infty} |a_{jk}|^2 \quad < \quad \infty \tag{J.28}$$

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 μ ($a_{jk} = a_{jk}(\mu)$, $j,k = 1,2,\cdots$) and both series in the defining condition converge uniformly in the domain of the parameter μ, 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 μ, over the domain of μ. (2) If the matrices $\mathbf{1} - \mathbf{A}$ and $\mathbf{1} - \mathbf{B}$ are von Koch matrices, then their product $\mathbf{1} - \mathbf{C} = (\mathbf{1} - \mathbf{A})(\mathbf{1} - \mathbf{B})$ is a von Koch matrix, and

$$\det(\mathbf{1} - \mathbf{C}) = \det(\mathbf{1} - \mathbf{A}) \det(\mathbf{1} - \mathbf{B}) \ . \tag{J.29}$$

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.28) 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.,

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1/2 & 0 & 0 & \cdots \\ 0 & 0 & 1/2 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 1/3 & \ddots \\ 0 & 0 & 0 & 0 & 1/3 & 0 & \ddots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix} \ . \tag{J.30}$$

Obviously, condition (J.27) is fulfilled by definition. Secondly, condition (J.28) 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.30) violates the trace-class definition (J.10), 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.5 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

$$R_n(z\mathbf{A}) = (\mathbf{1} + z\mathbf{A}) \exp\left(\sum_{k=1}^{n-1} \frac{(-z)^k}{k} \mathbf{A}^k\right) - \mathbf{1} \tag{J.31}$$

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

$$\det_n(\mathbf{1} + z\mathbf{A}) = \prod_{j=1}^{N(z\mathbf{A})} \left[(1 + z\lambda_j(\mathbf{A})) \exp\left(\sum_{k=1}^{n-1} \frac{(-z\lambda_j(\mathbf{A}))^k}{k}\right)\right] < \infty. \tag{J.32}$$

exists and is finite. The corresponding Plemelj-Smithies formula now reads [10]:

$$\det_n(\mathbf{1} + z\mathbf{A}) = \sum_{m=0}^{\infty} z^m \frac{\alpha_m^{(n)}(\mathbf{A})}{m!} \ . \tag{J.33}$$

with $\alpha_m^{(n)}(\mathbf{A})$ given by the $m \times m$ determinant:

$$\alpha_m^{(n)}(\mathbf{A}) = \begin{vmatrix} \sigma_1^{(n)} & m-1 & 0 & \cdots & 0 \\ \sigma_2^{(n)} & \sigma_1^{(n)} & m-2 & \cdots & 0 \\ \sigma_3^{(n)} & \sigma_2^{(n)} & \sigma_1^{(n)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & 1 \\ \sigma_m^{(n)} & \sigma_{m-1}^{(n)} & \sigma_{m-2}^{(n)} & \cdots & \sigma_1^{(n)} \end{vmatrix} \tag{J.34}$$

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.34) is that we get $\det_n(\mathbf{1}+\mathbf{A})$ from the standard Plemelj-Smithies formula (J.23) 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

$$\Delta(\lambda) = \prod_{k=0}^{\infty} \left(1 - \frac{\lambda}{\lambda_k}\right) \tag{J.35}$$

is regulated in the case $\mu \equiv d/m > 1$ as Weierstrass product

$$\Delta(\lambda) = \prod_{k=0}^{\infty} \left[\left(1 - \frac{\lambda}{\lambda_k}\right) \exp\left(\frac{\lambda}{\lambda_k} + \frac{\lambda^2}{2\lambda_k^2} + \cdots + \frac{\lambda^{[\mu]}}{[\mu]\lambda_k^{[\mu]}}\right)\right] \tag{J.36}$$

where $[\mu]$ denotes the integer part of μ. This is, see ref. [15], the unique entire function of order μ having zeros at $\{\lambda_k\}$ and subject to the normalization conditions

$$\ln\Delta(0) = \frac{d}{d\lambda}\ln\Delta(0) = \cdots = \frac{d^{[\mu]}}{d\lambda^{[\mu]}}\ln\Delta(0) = 0\ . \tag{J.37}$$

Clearly eq. (J.36) is the same as (J.32); 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

$$\Delta(E) = \det\left[(E - \mathbf{H})\right] \tag{J.38}$$

which – as it stands – would only make sense for a finite dimensional basis (or finite dimensional matrices). In ref. [17] 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

$$\Delta(E) = \det\left[(E - \mathbf{H})\Omega(E, \mathbf{H})\right] \tag{J.39}$$

where

$$\Omega(E, \mathbf{H}) = -\mathbf{H}^{-1} e^{E\mathbf{H}^{-1}} \tag{J.40}$$

such that the spectral determinant in the eigenbasis of $\mathbf{H}$ (with eigenvalues $E_n \neq 0$) reads

$$\Delta(E) = \prod_n \left(1 - \frac{E}{E_n}\right) e^{E/E_n} < \infty \ . \tag{J.41}$$

Note that $\mathbf{H}^{-1}$ is for this example of Hilbert-Schmidt character.

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 nonselfadjoint 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] A. Voros, *Unstable periodic orbits and semiclassical quantisation*, *J. Phys.* **A 21**, 685 (1988).

[J.17] J.P. Keating and M. Sieber, *Calculation of Spectral Determinants*, preprint 1994.

Appendix K

Solutions

Chapter 1

Solution 1.1: 3-disk symbolic dynamics. *Some of the cycles are listed in table 10.2 and drawn in fig. 1.4.*

Solution 11.5: 3-disk prime cycle counting. *The formula for arbitrary length cycles is derived in sect. 11.4.*

Solution 1.2: Sensitivity to initial conditions. *To estimate the pinball sensitivity we consider a narrow beam of point particles bouncing between two disks, fig. K.1(a). Or if you find this easier to visualize, think of a narrow ray of light. We assume that the ray of light is focused along the axis between the two points. This is where the least unstable periodic orbit lies, so its stability should give us an upper bound on the number of bounces we can expect to achieve. To estimate the stability we assume that the ray of light has a width $w(t)$ and a "dispersion angle" $\theta(t)$ (we assume both are small), fig. K.1(b). Between bounces*

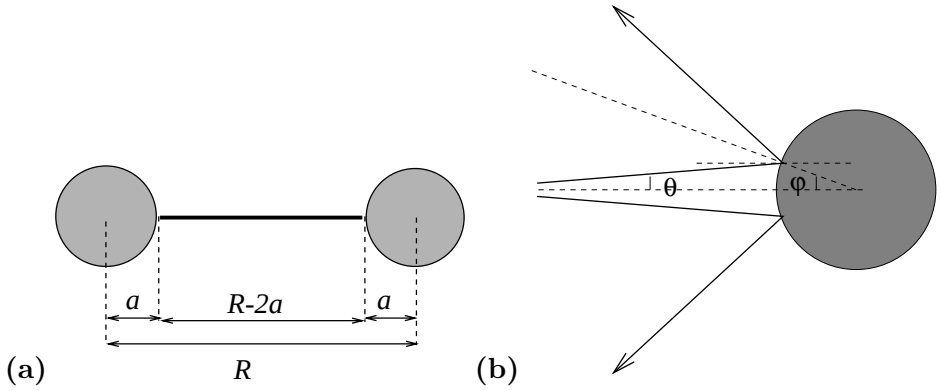


Figure K.1: The 2-disk pinball (a) geometry, (b) defocusing of scattered rays.

the dispersion angle stays constant while the width increases as

$$w(t) \approx w(t') + (t - t')\theta$$

At each bounce the width stays constant while the angle increases by

$$\theta_{n+1} = \theta_n + 2\phi \approx \theta_n + w(t)/a.$$

where θ_n denotes the angle after bounce n. Denoting the width of the ray at the nth bounce by w_n then we obtain the pair of coupled equations

$$w_{n+1} = w_n + \big(R - 2a\big)\,\theta_n \tag{K.1}$$

$$\theta_n = \theta_{n-1} + \frac{w_n}{a} \tag{K.2}$$

where we ignore corrections of order w_n^2 and θ_n^2. Solving for θ_n we find

$$\theta_n = \theta_0 + \frac{1}{a}\sum_{j=1}^{n} w_n.$$

Assuming $\theta_0 = 0$ then

$$w_{n+1} = w_n + \frac{R - 2a}{a}\sum_{j=1}^{n} w_n$$

Plugging in the values in the question we find the width at each bounce in Ångstrøms grows as 1, 5, 29, 169, 985, etc. To find the asymptotic behavior for a large number of bounces we try an solution of the form $w_n = ax^n$. Substituting this into the equation above and ignoring terms that do not grow exponentially we find solutions

$$w_n \approx a w_n^{asym} = a(3 \pm 2\sqrt{2})^n$$

The solution with the positive sign will clearly dominate. The constant a we cannot determine by this local analysis although it is clearly proportional to w_0. However, the asymptotic solution is a good approximation even for quite a small

number of bounces. To find an estimate of a we see that w_n/w_n^{asym} *very rapidly converges to 0.146447, thus*

$$w_n \approx 0.146447 w_0 (3+2\sqrt{2})^n \approx 0.1 \times w_0 \times 5.83^n$$

The outside edges of the ray of light will miss the disk when the width of the ray exceeds 2 cm; this occurs after 11 bounces.

(Adam Prügel-Bennett)

Solution 1.2: Sensitivity to initial conditions, another try. *Adam's estimate is not very good - do you have a better one? The first problem with it is that the instability is very underestimated. As we shall check in exercise 4.4, the exact formula for the 2-cycle stability is* $\Lambda = R - 1 + R\sqrt{1-2/R}$. *For* $R = 6$, $a = 1$ *this yields* $w_n/w_0 \approx (5+2\sqrt{6})^n = 9.898979^n$, *so if that were the whole story, the pinball would be not likely to make it much beyond 8 bounces.*

The second problem is that local instability overestimates the escape rate from an enclosure; trajectories are reinjected by scatterers. In the 3-disk pinball the particle leaving a disk can be reinjected by hitting either of other 2 disks, hence $w_n/w_0 \approx (9.9/2)^n$. *This interplay between local instability and global reinjection will be cast into the exact formula (**??**) involving "Lyapunov exponent" and "Kolmogorov entropy". In order to relate this estimate to our best continuous time escape rate estimate* $\gamma = 0.4103\ldots$ *(see table 13.2), we will have to also compute the mean free flight time (13.20). As a crude estimate, we take the shortest disk-to-disk distance,* $\langle T \rangle = R - 2 = 4$. *The continuous time escape rate result implies that* $w_n/w_0 \approx e^{(R-2)\gamma n} = (5.16)^n$, *in the same ballpark as the above expansion-reinjection estimate.*

(Predrag Cvitanović)

Solution 1.3: Trace-log of a matrix. *1) one method is to first check that this is true for any Hermitian matrix* M. *Then write an arbitrary complex matrix as sum* $M = A + zB$, A, B *Hermitian, Taylor expand in* z *and prove by analytic continuation that the identity applies to arbitrary* M.

(David Mermin)

2) another method: evaluate $\frac{d}{dt}\det\left(e^{t\ln M}\right)$ *by definition of derivative in terms of infinitesimals.*

(Kasper Juel Eriksen)

3) check appendix J.1

Chapter 2

Solution 2.2: Evolution as a group. *Can you see any other group replacing time? Try $\mathbb{Z}_{17}$ as an example; to mess things up try a non-commutative group.*

(Ronnie Mainieri)

Solution 2.11: Classical collinear helium dynamics. *An example of a solution are A. Prügel-Bennett's programs, available at* `www.nbi.dk/ChaosBook/extras/`.

Chapter 3

Solution 3.7: A pinball simulator. *An example of a pretty Xwindows pinball is A. Prügel-Bennett's xpinball.c program, available at* `www.nbi.dk/ChaosBook/extras/xpinball.tar.gz`.

Chapter 4

Solution 4.1: How unstable is the Hénon attractor?

1. *Evaluate numerically the Lyapunov exponent by iterating the Hénon map; For $a = 1.4$, $b = 0.3$ the answer should be close to $\lambda = 0.41$. If you have a good estimate and a plot of the convergence of your estimate with t, please send us your results for possibel inclusion into this text.*

2. *Both Lyapunov exponents for $a = 1.39945219$, $b = 0.3$ are negative, roughly $\lambda_1 = -0.2712, \lambda_2 = -0.9328$ (check that these values do respect the constant volume contraction condition (4.54) for the Hénon map). Why? because after a long transient exploration of the Hénon map's non–wandering set, on average after some 11,000 iterates, almost every inital point fall into a stable 13-cycle. You can check its existence by starting at on of its periodic points $(x_p, y_p = -0.2061, -0.3181)$.*

   If you missed the stable 13-cycle (as all students in one of the courses did), you should treat your computer experiments with great deal of scepticism.

As the product of eigenvalues is constant $-b$, you need to evaluate only the expanding eigenvalue. There are many ways to implement this calculation - here are a few:

1. *The most naive way - take the log of distance of two nearby trajectories, iterate until you run out of accuracy. Tray this many times, estimate an average.*

2. *slighly smarter: as above, but keep rescaling the length of the vector connecting neighboring points so it remains small, average over the sum of logs of rescaling factors. You can run this forever.*

3. *keep multiplying the [2×2] Jacobin stability matrix (4.53) until you run out of accuracy. Compute the log of the leading eigenvalue (4.21), tray this many times, estimate an average.*

4. *slighly smarter: as above, but start with an arbitrary inital tangent space vector, keep rescaling the length of the vector connecting neighboring points so it remains small. Can run this forever.*

5. *There is probably no need to use the QR decomposition method or any other such numerical method for this 2-dimensional problem.*

(Yueheng Lan and P. Cvitanović)

Solution 4.3: Billiard exercises. *Korsch and Jodl [?] have a whole book of numerical exercises with billiards, including 3-disks.*

Solution 2.7: Linearization for maps. *(difficulty: medium) The first few terms of the map h that conjugates f to αz*

$$f(z) = h^{-1}(\alpha h(z))\,.$$

are determined many places, for example in ref. [7].

There are conditions on the derivative of f at the origin to assure that the conjugation is always possible. These conditions are formulated in ref. [11], among others.

Chapter 5

Solution 5.1: Integrating over Dirac delta functions.

(b) It does not.

(c) Integrate by parts

$$\begin{aligned} 0 &= \int dx \frac{\partial}{\partial x} \left(g(x)\delta(f(x))\right) \\ &= \int dx \left(g'(x)\delta(f(x)) + g(x)f'(x)\delta'(f(x))\right) \end{aligned}$$

Taking $g(x) = 1/f'(x)$ we obtain

$$\int dx\, \delta'(f(x)) = \sum_{x^* \in Zero\, f} \frac{f''(x^*)}{|f'(x^*)|^3}$$

Chapter 8

Solution 8.1: Numerical estimate of the escape rate for a 1-d repeller *The logistic map is defined by $x_{n+1} = Ax_n(1-x_n)$. For $A \leq 4$ any point in the unit interval $[0,1]$ will remain in the interval forever. For $A > 4$ almost all points starting in the unit interval will eventually escape towards $-\infty$.*

The rate of escape can be easily measured by numerical experiment. We define the fraction of initial conditions that leave the interval after n iterations to be Γ_n. Fig. K.2 shows a plot of $\log(\Gamma_n)$ versus n, computed by starting with 10 000 000 random initial points. Asymptotically the escape rate falls off exponentially as

$$\Gamma(n) = C\mathrm{e}^{-\gamma n}\,.$$

Fig. K.2 suggests that this formula is very accurate even for relatively small n. We estimate γ by measuring the slope of the curve in fig. K.2. To avoid errors due to rounding and transients only the points $5 \leq n \leq 10$ were used. A linear regression fit yields the escape rate for $A = 6$:

$$\gamma = 0.8315 \pm 0.0001\,,$$

where the error is from statistical fluctuations (there may be systematic errors either due to rounding or because we are not in the true asymptotic regime).

(Adam Prügel-Bennet)

Solution 11.9: Spectrum of the "golden mean" pruned map.

1. *The idea is that with the redefinition 2 = 10, the alphabet {1,2} is unrestricted binary, and due to the piecewise linearity of the map, the stability weights factor in a way similar to (7.10).*

2. *As in (8.9), the spectral determinant for the Perron-Frobenius operator takes form (8.21)*

$$\det(1 - z\mathcal{L}) = \prod_{k=0}^{\infty} \frac{1}{\zeta_k}\,, \qquad \frac{1}{\zeta_k} = \prod_p \left(1 - \frac{z^{n_p}}{|\Lambda_p|\Lambda_p^k}\right)\,.$$

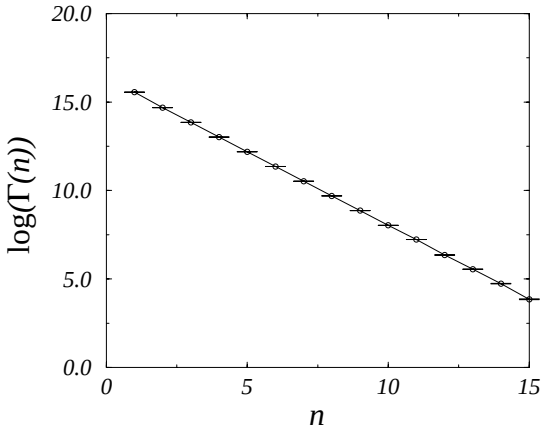


Figure K.2: Plot of $\log(\Gamma(n))$ versus n for the logistic map $x_{n+1} = 6x_n(1-x_n)$. Error bars show estimated errors in the mean assuming a binomial distribution. 10 000 000 random initial starting points were used.

The mapping is piecewise linear, so the form of the topological zeta function worked out in (11.15) already suggests the form of the answer. The alphabet {1,2} is unrestricted binary, so the dynamical zeta functions receive contributions only from the two fixed points, with all other cycle contributions cancelled exactly. The $1/\zeta_0$ is the spectral determinant for the transfer operator like the one in (5.13) with the $T_{00} = 0$, and in general

$$\begin{aligned} \frac{1}{\zeta_k} &= \left(1 - \frac{z}{|\Lambda_1|\Lambda_1^k}\right)\left(1 - \frac{z^2}{|\Lambda_2|\Lambda_2^k}\right)\left(1 - \frac{z^3}{|\Lambda_{12}|\Lambda_{12}^k}\right)\cdots \\ &= 1 - (-1)^k\left(\frac{z}{\Lambda^{k+1}} + \frac{z^2}{\Lambda^{2k+2}}\right) . \end{aligned} \tag{K.3}$$

The factor $(-1)^k$ arises because both stabilities Λ_1 and Λ_2 include a factor $-\Lambda$ from the right branch of the map. This answer contradicts (11.36). Which answer is the right one?

Solution 8.2: Dynamical zeta functions

1. *Work through section sect. 8.3.2.*

2. *Generalize the transition matrix (10.27) to a transfer operator.*

Solution 8.4: Dynamical zeta functions as ratios of spectral determinants. *Try inserting a factor equal to one in the zeta function and then expanding it. The problem is solved in sect. 8.5.1.*

Solution 13.3: Escape rate for the Ulam map. *The answer is worked out in* Nonlinearity **3**, *325;* **3**, *361 (1990).*

Solution 5.7: Eigenvalues of the skew Ulam tent map Perron-Frobenius operator. *The first few eigenvalues are*

$$
\begin{aligned}
e^{s_0} &= 1\,, \qquad e^{s_1} = \frac{2}{\Lambda_0} - 1 \\
e^{s_2} &= \frac{1}{4} + \frac{3}{4}\left(\frac{2}{\Lambda_0} - 1\right)^2\,, \qquad e^{s_3} = \frac{1}{2}\left(\frac{2}{\Lambda_0} - 1\right) + \frac{1}{2}\left(\frac{2}{\Lambda_0} - 1\right)^3 \ldots
\end{aligned}
$$

Solution 8.8: Dynamical zeta functions for Hamiltonian maps. *Read sect. 8.5.2.*

Solution 5.5: Invariant measure. *Compare the second map to the construction of Exercise 11.7.*

Solution 9.3: Euler formula. *Let*

$$
P = \prod_{k=0}^{\infty}(1 + tu^k) = \sum_{n=0}^{\infty} P_n t^n
$$

then

$$
\begin{aligned}
P_n &= \frac{1}{n!}\left.\frac{\partial^n P}{\partial t^n}\right|_{t=0} = \frac{1}{n!}\sum_{i_n \neq i_{n-1} \neq \cdots \neq i_1} u^{i_n + i_{n-1} + \cdots + i_1} \\
&= \sum_{i_n > i_{n-1} > \cdots i_1 \geq 0} u^{i_n + i_{n-1} + \cdots + i_1}
\end{aligned}
\tag{K.4}
$$

Clearly $P_0 = 1$, *and*

$$
P_1 = \sum_{i=0} u^i
$$

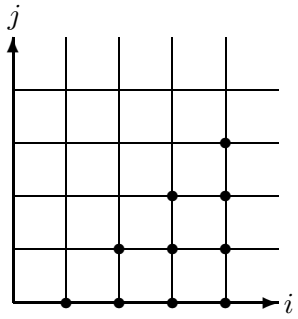
multiplying both sides by $1 - u$

$$
(1-u)P_1 = 1 + u + u^2 + \cdots - (u + u^2 + \cdots) = 1
$$

(since, for $|u| < 1$, $\lim_{n\to\infty} u^n = 0$). Thus $P_1 = 1/(1-u)$. Similarly

$$P_2 = \sum_{i>j\geq 0} u^{i+j}$$

Graphically the allowed values of i and j are



Performing the same trick as for P_1

$$(1-u)P_2 = \sum_{i>j\geq 0} u^{i+j} - \sum_{i>j\geq 0} u^{i+(j+1)}$$

The only terms that survive are those for which $j = i-1$ (that is the top diagonal in the figure) thus

$$(1-u)P_2 = u^{-1}\sum_{i=1}^{\infty} u^{2i}$$

and

$$(1-u)(1-u^2)P_2 = u^{-1}\left(u^2 + u^4 + \cdots - (u^4 + u^6 + \cdots)\right) = u$$

Thus

$$P_2 = \frac{u}{(1-u)(1-u^2)}$$

In general

$$(1-u)P_n = \sum_{i_n > i_{n-1} > \cdots i_1 \geq 0} u^{i_n + i_{n-1} + \cdots + i_1} - \sum_{i_n > i_{n-1} > \cdots i_1 \geq 0} u^{i_n + i_{n-1} + \cdots + (i_1+1)} \tag{K.5}$$

$$= u^{-1} \sum_{i_n > i_{n-1} > \cdots i_2 \geq 1} u^{i_n + i_{n-1} + \cdots + 2i_2} \tag{K.6}$$

since only the term $i_1 = i_2 - 1$ survives. Repeating this trick

$$(1-u)(1-u^2)P_n = u^{-1-2} \sum_{i_n > i_{n-1} > \cdots i_3 \geq 2} u^{i_n + i_{n-1} + \cdots + 3i_3}$$

and

$$\prod_{i=1}^{n} (1-u^i)\, P_n = u^{-(1+2+\cdots+n)} u^{n(n-1)} = u^{n(n-1)/2}$$

Thus

$$P_n = \frac{u^{n(n-1)/2}}{\prod_{i=1}^{n}(1-u^i)}.$$

(Adam Prügel-Bennet)

Solution 9.3: Euler formula, 2nd method. *The coefficients Q_k in (9.3) are given explicitly by the* Euler formula

$$Q_k = \frac{1}{1-\Lambda^{-1}} \, \frac{\Lambda^{-1}}{1-\Lambda^{-2}} \, \cdots \, \frac{\Lambda^{-k+1}}{1-\Lambda^{-k}} \quad . \tag{K.7}$$

Such a formula is easily proved by considering the finite order product

$$\mathcal{W}_j(z,\gamma) \;=\; \prod_{l=0}^{j} (1 + z\gamma^l) \;=\; \sum_{l=0}^{j+1} \Gamma_l z^l$$

Since we have that

$$(1 + z\gamma^{j+1})\mathcal{W}_j(z,\gamma) \;=\; (1+z)\mathcal{W}_j(\gamma z, \gamma)\,,$$

we get the following identity for the coefficients

$$\Gamma_m + \Gamma_{m-1}\gamma^{j+1} \;=\; \Gamma_m \gamma^m + \Gamma_{m-1}\gamma^{m-1} \quad m = 1, \ldots \,.$$

Starting with $\Gamma_0 = 1$, we recursively get

$$\Gamma_1 \;=\; \frac{1-\gamma^{j+1}}{1-\gamma} \qquad \Gamma_2 \;=\; \frac{(1-\gamma^{j+1})(\gamma-\gamma^{j+1})}{(1-\gamma)(1-\gamma^2)} \cdots \,.$$

the Euler formula (9.4) follows once we take the $j \to \infty$ limit for $|\gamma| < 1$.

(Robert Artuso)

Solution 9.3: Euler formula, 3rd method. *First define*

$$f(t,u) := \prod_{k=0}^{\infty}(1+tu^k) \ . \tag{K.8}$$

Note that

$$f(t,u) = (1+t)f(tu,u) \ , \tag{K.9}$$

by factoring out the first term in the product. Now make the ansatz

$$f(t,u) = \sum_{n=0}^{\infty} t^n g_n(u) \ , \tag{K.10}$$

plug it into (K.9), compare the coefficients of t^n and get

$$g_n(u) = u^n g_n(u) + u^{n-1} g_{n-1}(u) \ . \tag{K.11}$$

Of course $g_0(u) = 1$. Therefore by solving the recursion (K.11) and by noting that $\sum_{k=1}^{n-1} k = \frac{n(n-1)}{2}$ one finally arrives at

$$g_n(u) = \frac{u^{\frac{n(n-1)}{2}}}{\prod_{k=1}^{n}(1-u^k)} \ . \tag{K.12}$$

Euler got this formula and he and Jacobi got many nice number theoretical results from it, most prominent the pentagonal number theorem, which says that in the series expansion of $\prod_{k=1}^{\infty}(1-q^k)$ all terms cancel except those which have as an exponent the circumference of a regular pentagon with integer base length.

(Juri Rolf)

Solution 9.4: 2-d product expansion. *Now let us try to apply the same trick as above to the two dimensional situation*

$$h(t,u) := \prod_{k=0}^{\infty}(1+tu^k)^{k+1} \ . \tag{K.13}$$

Write down the first terms and note that similar to (K.9)

$$h(t,u) = f(t,u)h(tu,u) \ , \tag{K.14}$$

where f is the Euler product (K.8). Now make the ansatz

$$h(t,u) = \sum_{n=0}^{\infty} t^n a_n(u) \tag{K.15}$$

and use the series expansion for f in (K.14) to get the recursion

$$a_n(u) = \frac{1}{1-u^n} \sum_{m=0}^{n-1} u^m a_m(u) g_{n-m}(u) \ . \tag{K.16}$$

With this one can at least compute the generalized Euler product effectively, but it would be nice if one could use it for a proof of the general behaviour of the coefficients a_n.

(Juri Rolf)

Chapter 10

Solution 10.4: Reduction of 3-disk symbolic dynamics. *The answer is given in sect. 17.6.*

Solution 10.5: Unimodal map symbolic dynamics. *Hint: write down an arbitrary binary number such as $\gamma = .1101001101000\ldots$ and generate the future itinerary S^+ by checking whether $f^n(\gamma)$ is greater or less than 1/2. Then verify that (10.18) recovers γ.*

Chapter 11

Solution 11.4: Transition matrix and cycle counting.

e) The topological entropy is

$$h = \ln\left(1+\frac{\sqrt{5}}{2}\right) \ .$$

Solution 11.44: Alphabet {0,1}, prune _1000_, _00100_, _01100_.

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. Physically 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 exercise 11.20, 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_. Physically, 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 exercise 11.8, 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. The topological zeta function is given by*

$$1/\zeta = (1-t_0)(1-t_1-t_2-t_{23}-t_{113}) \tag{K.17}$$

for unrestricted 4-letter alphabet {1, 2, $\underline{23}$, $\underline{113}$}.

Solution 11.13: Whence Möbius function? *Written out $f(n)$ line-by-line for a few values of n, (11.37) yields*

$$\begin{aligned} f(1) &= g(1) \\ f(2) &= g(2)+g(1) \\ f(3) &= g(3)+g(1) \\ f(4) &= g(4)+g(2)+g(1) \\ &\cdots \\ f(6) &= g(6)+g(3)+g(2)+g(1) \\ &\cdots \end{aligned} \tag{K.18}$$

Now invert recursively this infinite tower of equations to obtain

$$\begin{aligned} g(1) &= f(1) \\ g(2) &= f(2)-f(1) \\ g(3) &= f(3)-f(1) \end{aligned}$$

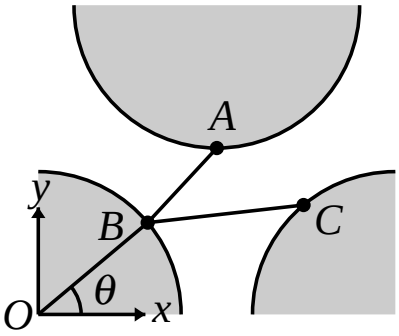


Figure K.3: Minimizing the path from the previous bounce to the next bounce.

$$
\begin{aligned}
g(4) &= f(4) - [f(2) - f(1)] - f(1) = f(4) - f(2) \\
&\dots \\
g(6) &= f(6) - [f(3) - f(1)] - [f(2) - f(1)] - f(1) \\
&\dots
\end{aligned}
$$

We see that $f(n)$ contributes with factor -1 if n prime, and not at all if n contains a prime factor to a higher power. This is precisely the raison d'etre for the Möbius function, with whose help the inverse of (11.37) can be written as the Möbius inversion formula [?] (11.38).

Chapter 12

Solution 12.3: Stability of billiard cycles. *The 2-cycle $\overline{0}$ stability (4.4) is the solution to both problems (provided you evaluate correctly the hyperbola curvature on the diagonal).*

Solution 12.4: Numerical cycle routines. *A number of sample Fortran programs for finding periodic orbits is available on the homepage for this manuscript,* `www.nbi.dk/ChaosBook/`.

Solution 12.9: Billiard cycles by path length minimization *in a given order is to start with a guess path where each bounce is given some arbitrary position on the correct disk and then iteratively improve on the guess. To accomplish this an improvement cycle is constructed whereby each bouncing point in the orbit is taken in turn and placed in a new position so that it minimizes the path. Since the positions of all the other bounces are kept constant this involves choosing the new bounce position which minimizes the path from the previous bounce to the next bounce. This problem is schematically represented in fig. K.3*

Finding the point B involves a one dimensional minimization. We define the vectors $\vec{A} = \vec{OA}$, $\vec{B} = \vec{OB}$ and $\vec{C} = \vec{OC}$. We wish to minimize the length L_{ABC}

by varying $\vec{B}$ subject to the constraint that $|\vec{B}| = a$. Clearly

$$\begin{aligned} L_{ABC} &= \left|\vec{A} - \vec{B}\right| + \left|\vec{C} - \vec{B}\right| \\ &= \sqrt{\vec{A}^2 + \vec{B}^2 - 2\vec{A}\cdot\vec{B}} + \sqrt{\vec{C}^2 + \vec{B}^2 - 2\vec{C}\cdot\vec{B}} \end{aligned}$$

writing

$$\vec{B}(\theta) = a(\cos\theta, \sin\theta)$$

then the minima is given by

$$\frac{dL_{ABC}}{d\theta} = -\left(\frac{\vec{A}}{\sqrt{\vec{A}^2 + \vec{B}^2 - 2\vec{A}\cdot\vec{B}}} + \frac{\vec{C}}{\sqrt{\vec{C}^2 + \vec{B}^2 - 2\vec{C}\cdot\vec{B}}}\right)\cdot\vec{B}'(\theta) = 0.$$

The minima can then be found using a bisection algorithm or using Newton-Raphson. A simpler way is to observe that $\vec{B}'(\theta)$ is orthogonal to $\vec{B}(\theta)$ so that the vector

$$\vec{D} = \frac{\vec{A}}{\sqrt{\vec{A}^2 + \vec{B}^2 - 2\vec{A}\cdot\vec{B}}} + \frac{\vec{C}}{\sqrt{\vec{C}^2 + \vec{B}^2 - 2\vec{C}\cdot\vec{B}}}$$

will be proportional to $\vec{B}$. This then provides an iterative sequence for finding $\vec{B}$

- *Starting from your current guess for $\vec{B}$ calculate $\vec{D}$*
- *Put $\vec{B} = a\vec{D}/|\vec{D}|$*
- *Repeat the first step until you converge.*

At each iteration of the improvement cycle the total length of the orbit is measured. The minimization is complete when the path length stops improving. Although this algorithm is not as fast as the Newton-Raphson method, it nevertheless converges very rapidly.

(Adam Prügel-Bennet)

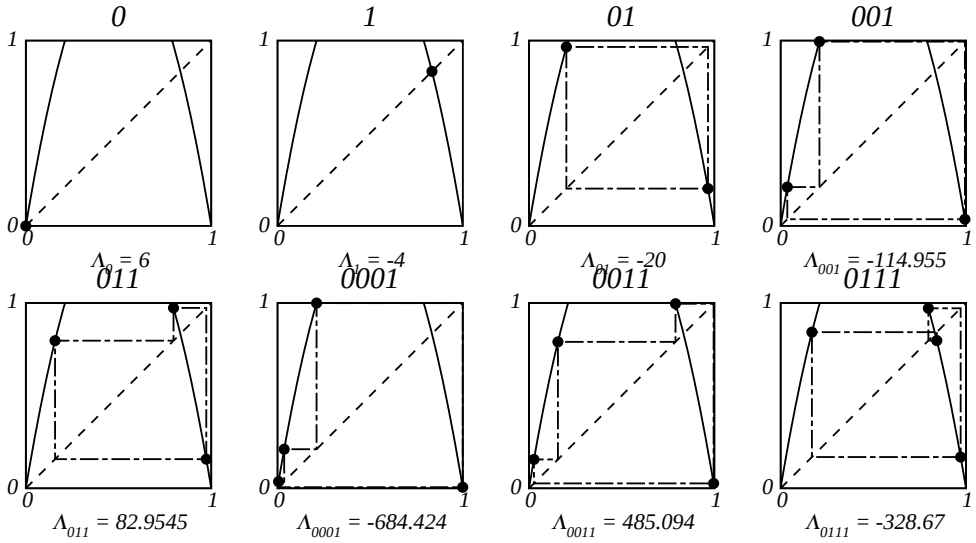


Figure K.4: Periodic orbits and stabilities for the logistics equation $x_{n+1} = 6x_n(1 - x_n)$.

Chapter 13

Solution 13.2: Prime cycles for a 1-d repeller, analytic fromulas. *For the logistic map the prime cycles, ordered in terms of their symbolic dynamics, are listed in table 10.1*

$$\mathcal{P} = \{0, 1, 01, 001, 011, 0001, 0011, 0111, \ldots\}$$

The position of the prime cycles can be found by iterating the inverse mapping. If we wish to find the position of a prime orbit $p = b_1 b_2 \cdots b_{n_p}$, where $b_i \in \{0, 1\}$, then starting from some initial point, $x = 1/2$ say, we apply one of the inverse mappings

$$f_{\pm}^{-1}(x) = \frac{1}{2} \pm \frac{1}{2}\sqrt{1 - x/4A}$$

where we choose f_-^{-1} if $b_1 = 0$ or f_+^{-1} if $b_1 = 1$. We then apply the inverse mapping again depending on the next element in the prime orbit. Repeating this procedure many times we converge onto the prime cycle. The stability Λ_p of a prime cycle p is given by the product of slopes of f around the cycle. The first eight prime cycles are shown in fig. K.4.

The stabilities of the first five prime orbits can be calculated for arbitrary A. We find that $\Lambda_0 = A$, $\Lambda_1 = 2 - A$, $\Lambda_{01} = 4 + 2A - A^2$, and

$$\Lambda_{\substack{001\\011}} = 8 + 2A - A^2 \pm A(2 - A)\sqrt{A^2 - 2A - 7}. \tag{K.19}$$

There is probably a closed form expression for the 4-cycles as well.

For crosschecking purposes: if $A = 9/2$, $\Lambda_0 = 9/2$ $\Lambda_1 = -5/2$ $\Lambda_{01} = -7.25$ $\Lambda_{011} = 19.942461\ldots$.

(Adam Prügel-Bennet)

Solution 13.2: Dynamical zeta function for a 1-d repeller *The escape rate can be estimated from the leading zero in the dynamical zeta function* $1/\zeta(z)$, *defined by*

$$1/\zeta(z) = \prod_p (1 - z^{n_p}/|\Lambda_p|) \ .$$

To compute the position of this pole we expand $1/\zeta(z)$ *as a power series (13.5) in* z

$$1/\zeta(z) = 1 - \sum_{i=1} \hat{c}_i z^i$$

where

$$\begin{aligned} \hat{c}_1 &= |\Lambda_0|^{-1} + |\Lambda_1|^{-1}\,, \qquad \hat{c}_2 = |\Lambda_{01}|^{-1} - |\Lambda_1\Lambda_0|^{-1} \\ \hat{c}_3 &= |\Lambda_{001}|^{-1} - |\Lambda_0\Lambda_{01}|^{-1} + |\Lambda_{011}|^{-1} - |\Lambda_{01}\Lambda_1|^{-1} \end{aligned}$$

etc.. Using the cycles up to length 6 we get

$$\begin{aligned} 1/\zeta(z) &= 1 - 0.416667z - 0.00833333z^2 \\ &\quad +0.000079446z^3 - 9.89291\times 10^{-7} z^4 + \ldots \end{aligned}$$

The leading zero of this Taylor series is an estimate of $\exp(\gamma)$. *Using* $n = 1$, *2, 3 and 4 we obtain the increasingly accurate estimates for* γ: *0.875469, 0.830597, 0.831519 and 0.831492 In a hope to improve the convergence we can use the Padé approximates* $P_M^N(z) = \sum_{i=1}^N p_i z^i/(1 + \sum_{j=1}^M q_j z^j)$. *Using the Padé approximates* $P_1^{n-1}(z)$ *for* $n = 2$, *3 and 4 we obtain the estimates 0.828585, 0.831499 and 0.831493.*

The above results correspond to $A = 6$; *in the* $A = 9/2$ *case the leading zero is* $1/z = 1.43549\ldots$ *and* $\gamma = 0.36150\ldots$. *(Adam Prügel-Bennet)*

Solution 13.2: Spectral determinant for a 1-d repeller *We are told the correct expression for the escape rate is also given by the logarithm of the*

leading zero of the spectral determinant (8.21), expanded as the Taylor series (13.8). The coefficients c_i should fall off super-exponentially so that truncating the Taylor series is expected to give a far more accurate estimate of the escape rate than using the dynamical zeta function. How do we compute the c_i coefficients in (13.8)? One straightforward method is to first compute the Taylor expansion of $\log(F(z))$

$$\begin{aligned}\log(F(z)) &= \sum_p \sum_{k=0} \log\left(1-\frac{t_p}{\Lambda_p^k}\right) = -\sum_p \sum_{k=0} \sum_{r=1} \frac{t_p^r}{\Lambda_p^{kr}} \\ &= -\sum_p \sum_{r=1} \frac{t_p^r}{1-\Lambda_p^{-r}} = -\sum_p \sum_{r=1} B_p(r) z^{n_p r}\end{aligned}$$

where $B_p(r) = -1/r|\Lambda_p^r|(1+\Lambda_p^{-r})$. *Writing* $\log(F(z))$ *as a power series*

$$\log(F(z)) = -\sum_{i=1} b_i z^i$$

we obtain

$$\begin{aligned}b_1 &= B_0(1) + B_1(1) \\ b_2 &= B_{01}(1) + B_0(2) + B_1(2) \\ b_3 &= B_{001}(1) + B_{011}(1) + B_0(3) + B_1(3) \\ b_3 &= B_{0001}(1) + B_{0011}(1) + B_{0111}(1) + B_{01}(2) + B_0(4) + B_1(4) \qquad \text{(K.20)}\end{aligned}$$

etc.. To obtain the coefficients for the spectral determinant we solve

$$F(z) = 1 - \sum_{i=1} Q_i z^i = \exp\left(\sum_{i=1} b_i z^i\right)$$

for the Q_i's. This gives

$$\begin{aligned}Q_1 &= b_1\,, \quad Q_2 = b_2 + b_1^2/2\,, \quad Q_3 = b_3 + b_1 b_2 + b_1^3/6 \\ Q_4 &= b_4 + b_1 b_3 + b_2^2/2 + b_2 b_1^2/2 + b_1^4/24\end{aligned}$$

Using these formulas we find

$$F(z) = 1 - 0.4z - 0.0152381z^2 - 0.0000759784z^3 + 4.5311 \times 10^{-9} z^4 + \cdots$$

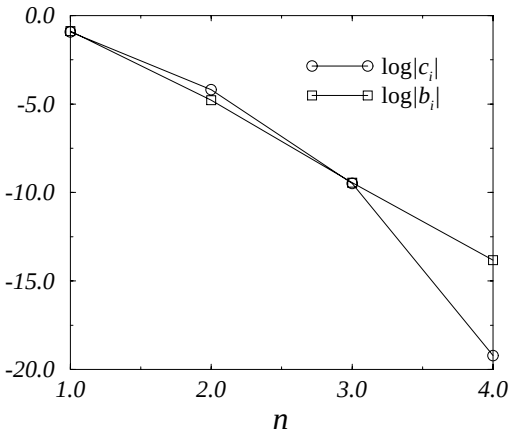


Figure K.5: Plot of the Taylor coefficients for the spectral determinant, c_i, and for the dynamical zeta function, b_i.

The logarithm of the leading zero of $F(z)$ again gives the escape rate. Using the $n = 1$, 2, 3, and 4 truncations we find the approximation to γ of 0.916291, 0.832345, 0.83149289 and 0.8314929875. As predicted, the convergence is much faster for the spectral determinant than for the dynamical zeta function.

In fig. K.5 we show a plot of the logarithm of the coefficients for the spectral determinant and for the dynamical zeta function.

(Adam Prügel-Bennet)

The above results correspond to $A = 6$; in the $A = 9/2$ case all cycles up to length 10 yield $\gamma = 0.36150966984250926\ldots$. *(Vadim Moroz)*

Solution 13.2: Functional dependence of escape rate for a 1-d repeller *We can compute an approximate functional dependence of the escape rate on the parameter a using the stabilities of the first five prime orbits computed above, see (K.19). The spectral determinant (for $a > 4$) is*

$$\begin{aligned} F &= 1 - \frac{2z}{a-1} - \frac{8z^2}{(a-3)(a-1)^2(a+1)} \\ &\quad + \left(\frac{2(32 - 18a + 17a^2 - 16a^3 + 14a^4 - 6a^5 + a^6)}{(a-3)(a-1)^3(1+a)(a^2-5a+7)(a^2+a+1)} \right. \\ &\quad \left. - \frac{2a(a-2)\sqrt{(a^2-2a-7)}}{(a^2-5a+7)(a^2-2a-7)(a^2+a+1)} \right) z^3 \end{aligned} \tag{K.21}$$

The leading zero is plotted in fig. K.6; it always remains real while the other two roots which are large and negative for $a > 5.13\ldots$ become imaginary below this

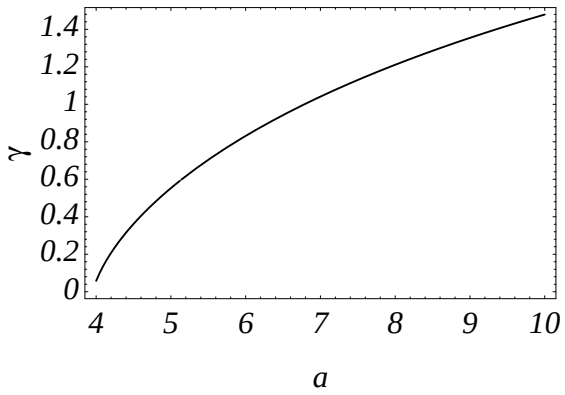


Figure K.6: Plot of the escape rate versus a for the logistic map $x_{n+1} = ax_n(1 - x_n)$ calculated from the first five periodic orbits.

critical value. The accuracy of this truncation is clearly worst for $a \to 4$, the value at which the hyperbolicity is lost and the escape rate goes to zero.

(Adam Prügel-Bennet)

Solution 13.3: Escape rate for the Ulam map. *The answer is given in ref. [2].*

Chapter 14

Solution 5.4: *The escape rate is the leading zero of the zeta function*

$$0 = 1/\zeta(\gamma) = 1 - e^{\gamma}/2a - e^{\gamma}/2a = 1 - e^{\gamma}/a.$$

So, $\gamma = \log(a)$ if $a > a_c = 1$ and $\gamma = 0$ otherwise. For $a \approx a_c$ the escape rate behaves like

$$\gamma(a) \approx (a - a_c).$$

Solution 14.1: *The escape is controlled by the size of the primary hole of the repeller. All subholes in the repeller will be proportional with the main hole. The size of the main hole is $l = \sqrt{1 - 1/a}$. Near $a_c = 1$ the escape rate is*

$$\gamma(a) \sim (a - a_c)^{1/2}.$$

We can generalize this and the previous result and conclude that

$$\gamma(a) \sim (a - a_c)^{1/z},$$

where z is the order of the maximum of the single humped map.

Solution 14.2: *By direct evaluation we can calculate the zeta functions and the Fredholm determinant of this map. The zeta functions are*

$$1/\zeta_k(z) = \det(1 - z\mathbf{T}_k),$$

where

$$\mathbf{T}_k = \begin{pmatrix} T_{00}^{k+1} & T_{01}^{k+1} \\ T_{10}^{k+1} & T_{11}^{k+1} \end{pmatrix},$$

and $T_{00} = 1/a_1$, $T_{01} = (b - b/a_1)/(1-b)$, $T_{11} = (1-b-b/a_2)/(1-b)$, $T_{10} = 1/a_2$ are inverses of the slopes of the map. The Fredholm determinant is the product of zeta functions

$$F(z) = \prod_{k=0}^{\infty} 1/\zeta_k(z).$$

The leading zeroes of the Fredholm determinant can come from the zeroes of the leading zeta functions.

The zeroes of $1/\zeta_0(z)$ are

$$\begin{aligned} 1/z_1 &= \frac{T_{00}+T_{11}+\sqrt{(T_{00}-T_{11})^2+4T_{01}T_{10}}}{2}, \\ 1/z_2 &= \frac{T_{00}+T_{11}-\sqrt{(T_{00}-T_{11})^2+4T_{01}T_{10}}}{2}. \end{aligned}$$

The zeroes of $1/\zeta_1(z)$ are

$$\begin{aligned} 1/z_3 &= \frac{T_{00}^2+T_{11}^2+\sqrt{(T_{00}^2-T_{11}^2)^2+4T_{01}^2T_{10}^2}}{2}, \\ 1/z_4 &= \frac{T_{00}^2+T_{11}^2-\sqrt{(T_{00}^2-T_{11}^2)^2+4T_{01}^2T_{10}^2}}{2}. \end{aligned}$$

By substituting the slopes we can show that $z_1 = 1$ is the leading eigenvalue. The next to leading eigenvalue, which is the correlation decay in discrete time, can be $1/z_3$ or $1/z_2$.

Chapter 15

Solution 15.1: *In the higher dimensional case there is no change in the derivation except Λ_p should be replaced with the product of expanding eigenvalues $\prod_j |\Lambda_{p,j}|$. The logarithm of this product is $\sum_j \log|\Lambda_{p,j}|$. The average of $\log|\Lambda_{,j}|$ is the j-th Lyapunov exponent.*

Solution 15.4: *The zeta function for the two scale map is*

$$1/\zeta(z,\beta) = 1 - z\left(\frac{1}{a^\beta} + \frac{1}{b^\beta}\right).$$

The pressure function is

$$P(\beta) = \log z_0(\beta) = -\log\left(\frac{1}{a^\beta} + \frac{1}{b^\beta}\right).$$

The escape rate is

$$\gamma = P(1) = -\log\left(\frac{1}{a} + \frac{1}{b}\right),$$

The topological entropy is

$$K_0 = h_{top} = -P(0) = \log 2.$$

The Lyapunov exponent is

$$\overline{\lambda} = P'(1) = \frac{\log a/a + \log b/b}{1/a + 1/b}.$$

The Kolmogorov entropy is

$$K_1 = \overline{\lambda} - \gamma = P'(1) - P(1) = \frac{\log a/a + \log b/b}{1/a + 1/b} + \log\left(\frac{1}{a} + \frac{1}{b}\right).$$

The Rényi entropies are

$$K_\beta = (P(\beta) - \beta\gamma)/(\beta - 1) = (\log\left(\frac{1}{a^\beta} + \frac{1}{b^\beta}\right) + \beta\log\left(\frac{1}{a} + \frac{1}{b}\right))/(1-\beta).$$

The box counting dimension is the solution of the implicit equation $P(D_0) = 0$, *which is*

$$1 = \frac{1}{a_0^D} + \frac{1}{b_0^D}.$$

The information dimension is

$$D_1 = 1 - \gamma/\overline{\lambda}.$$

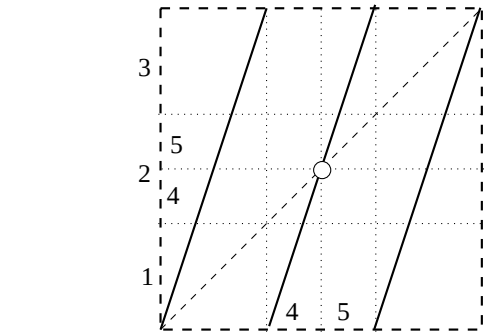
The rest of the dimensions can be determined from equation $P(q-(q-1)D_q) = \gamma q$. *Taking* exp *of both sides we get*

$$\frac{1}{a^{q-(q-1)D_q}} + \frac{1}{b^{q-(q-1)D_q}} = \left(\frac{1}{a} + \frac{1}{b}\right)^q.$$

For a given q *we can find* D_q *from this implicit equation.*

Solution 15.5: *The zeta function is*

$$1/\zeta(z,\beta) = \det\left(1 - \mathbf{T}_{\beta-1}\right),$$



(a) **(b)** **(c)** **(d)**

Figure K.7: (a) (b) A partition of the unit interval into three or five intervals, labeled by the order along the unit interval $\mathcal{A} = \{\mathcal{M}_1, \mathcal{M}_2 = \mathcal{M}_4 \cup (\frac{1}{2}) \cup \mathcal{M}_5, \mathcal{M}_3\}$. The partition is Markov, as the critical point is also a fixed point. (c) the Markov graph for this Markov partition.

where we replaced k with $\beta - 1$ in Solution K. The pressure can be calculated from the leading zero which is (see Solution K)

$$P(\beta) = \log z_0(\beta) = -\log\left(\frac{T_{00}^\beta + T_{11}^\beta + \sqrt{(T_{00}^\beta - T_{11}^\beta)^2 + 4T_{01}^\beta T_{10}^\beta}}{2}\right).$$

Solution 15.6: *We can easily read off that $b = 1/2$, $a_1 = arcsin(1/2)/2\pi$ and $a_2 = a_1$ and do the steps as before.*

Chapter 18

Solution 18.1: Diffusion for odd integer Λ. *Consider first the case $\Lambda = 3$, illustrated in fig. K.7. If $\beta = 0$, the dynamics in the elementary cell is simple enough; a partition can be constructed from three intervals, which we label $\{\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3\}$, with the alphabet ordered as the intervals are laid out along the unit interval. The Markov graph is fig. K.7(c), and the dynamical zeta function is*

$$1/\zeta|_{\beta=0} = 1 - (t_1 + t_2 + t_3) = 1 - 3z/\Lambda\,,$$

with eigenvalue $z = 1$ as required by the flow conservation.

However, description of global diffusion requires more care. As explained in the definition of the map (18.9), we have to split the partition $\mathcal{M}_2 = \mathcal{M}_4 \cup (\frac{1}{2}) \cup$

$\mathcal{M}_5$, and exclude the fixed point $f(\frac{1}{2}) = \frac{1}{2}$, as the map $\hat{f}(\hat{x})$ is not defined at $\hat{f}(\frac{1}{2})$. (Are we to jump to the right or to the left at that point?) As we have $f(\mathcal{M}_4) = \mathcal{M}_1 \cup \mathcal{M}_4$, and similarly for $f(\mathcal{M}_5)$, the Markov graph fig. K.7(d) is infinite, and so is the dynamical zeta function:

$$1/\zeta = 1 - t_1 - t_{14} - t_{144} - t_{1444} \cdots - t_3 - t_{35} - t_{355} - t_{3555} \cdots .$$

The infinite alphabet $\mathcal{A} = \{1, 14, 144, 1444 \cdots 3, 35, 355, 3555 \cdots\}$ is a consequence of the exclusion of the fixed point(s) x_4, x_5. As is customary in such situations (see sect. ?? exercise 13.10, and chapter 16, inter alia*), we deal with this by dividing out the undesired fixed point from the dynamical zeta function. We can factorize and resum the weights using the piecewise linearity of (18.9)*

$$1/\zeta = 1 - \frac{t_1}{1-t_4} - \frac{t_3}{1-t_5} .$$

The diffusion constant is now most conveniently evaluated by evaluating the partial derivatives of $1/\zeta$ as in (13.16)

$$\begin{aligned} \langle T \rangle_\zeta &= -z\frac{\partial}{\partial z}\frac{1}{\zeta} = 2\left(\frac{t_1}{1-t_4} + \frac{t_1 t_4}{(1-t_4)^2}\right)\Big|_{z=1,\beta=0} = \frac{3}{4} \\ \langle \hat{x}^2 \rangle_\zeta\Big|_{z=1,\beta=0} &= 2\left(\frac{\hat{n}_1(\hat{n}_1+\hat{n}_4)\Lambda^2}{(1-1/\Lambda)^2} + 2\frac{\hat{n}_4^2/\Lambda^3}{(1-1/\Lambda)^3}\right) = \frac{1}{2} \end{aligned} \tag{K.22}$$

yielding $D = 1/3$, in agreement with in (18.20) for $\Lambda = 3$.

Chapter 21

Solution 21.1: Lorentzian representation of the Dirac delta function. *To see that (21.15) is a delta function, express explicitely the imaginary part:*

$$-\lim_{\epsilon\to+0}\frac{1}{\pi}\mathrm{Im}\,\frac{E-E_n-i\epsilon}{(E-E_n)^2+\epsilon^2} = \lim_{\epsilon\to+0}\frac{1}{\pi}\frac{\epsilon}{(E-E_n)^2+\epsilon^2}. \tag{K.23}$$

This is a Lorenzian of width ϵ with a peak at $E = E_n$. It has a correct normalization for the delta function as

$$\frac{1}{\pi}\int_{-\infty}^{\infty} dE \frac{\epsilon}{(E-E_n)^2+\epsilon^2} = 1, \tag{K.24}$$

independently of the value of ϵ. Argue that in the $\epsilon \to \infty$ limit the support of the Lorentzian is concentrated at $E = E_n$, and providing that the function integrated over has a finite first derivative at $E = E_n$ and falls of sufficiently rapidly as $E \to \pm\infty$, this is a representation of the delta function.

Solution 21.6: Free particle action.

a) a d-dimensional free particle:
Lagrangian is the difference of kinetic and potential energy. For free motion the potential energy is $U(q) = 0$ and the velocity and the kinetic energy $m\dot{q}^2/2$ are constant. Integrating $\int_{t_0}^{t} d\tau L(q(\tau), \dot{q}(\tau), \tau)$ we obtain

$$R(q, q', t) = m(q - q')^2/2t$$

Solution 21.14: Stationary phase approximation. *The main contribution to such integrals comes from neighborhoods of values of x of stationary phase, the points for which the gradient of the phase vanishes*

$$\frac{\partial}{\partial x}\Phi(x) = 0.$$

Intuitively, these are the important contributions as for $\hbar \to 0$ the phase $\Phi(x)/\hbar$ grows large and the function $e^{i\Phi(x)/\hbar}$ oscillates rapidly as a function of x, with the negative and positive parts cancelling each other. More precisely, if the stationary points are well separated local extrema of $\Phi(x)$, we can deform the integration contour and approximate $\Phi(x)/\hbar$ up to the second order in x by

$$I \approx \sum_n A(x_n) e^{i\Phi(x_n)/\hbar} \int d^d x e^{\frac{i}{2\hbar}(x-x_n)^T \mathbf{D^2}\Phi(x_n)(x-x_n)}.$$

The second derivative matrix is a real symmetric matrix, so we can transform it to a diagonal matrix by a similarity transformation

$$Diag(\lambda_1, ..., \lambda_d) = \mathbf{O}\mathbf{D}^2\Phi\mathbf{O}^+ ,$$

where $\mathbf{O}$ is a matrix of an orthogonal transfomation. In the rotated coordinate system $u = \mathbf{O}(x - x_n)$ and the integral takes form

$$I \approx \sum_n A(x_n) e^{i\Phi(x_n)/\hbar} \int d^d u e^{\sum_{k=1}^{d} i\lambda_k u_k^2/2\hbar} ,$$

where we used the fact that the Jacobi determinant of an orthogonal transformation is $\det \mathbf{O} = 1$. *Carrying out the Gauss integrals*

$$\int du e^{i\lambda u^2/2\hbar} = \frac{(2\pi i\hbar)^{1/2}}{\sqrt{\lambda}} \tag{K.25}$$

and using $\det \mathbf{D}^2\Phi(x_n) = \prod_{k=1}^{d} \lambda_k$ *we obtain the stationary phase estimate of (21.73).*

A nice exposition of the subject is given in ref. [17].

Solution 21.18: A usefull determinant identity. *Divide out* E *in the last column of 21.75 and get the following matrix*

$$E \begin{bmatrix} x_{1,1} & \dots & x_{1,n} & y_1 E^{-1} \\ \vdots & \ddots & \vdots & \vdots \\ x_{n,1} & \dots & x_{n,n} & y_n E^{-1} \\ z_1 & \dots & z_n & 1 \end{bmatrix}$$

Now we subtract the last column multiplied with z_n *from the second last column (these matrix operations does not change the determinant) to get*

$$E \begin{bmatrix} x_{1,1} & \dots & x_{1,n-1} & x_{1,n} - z_n y_1 E^{-1} & y_1 E^{-1} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{n,1} & \dots & x_{n,n-1} & x_{n,n} - z_n y_n E^{-1} & y_n E^{-1} \\ z_1 & \dots & z_{n-1} & 0 & 1 \end{bmatrix}$$

This continues eliminating all the z_i*'s in the bottom row getting the following matrix*

$$E \begin{bmatrix} x_{1,1} - z_1 y_1 E^{-1} & \dots & x_{1,n} - z_n y_1 E^{-1} & y_1 E^{-1} \\ \vdots & \ddots & \vdots & \vdots \\ x_{n,1} - z_1 y_n E^{-1} & \dots & x_{n,n} - z_n y_n E^{-1} & y_n E^{-1} \\ 0 & \dots & 0 & 1 \end{bmatrix}$$

and we get (21.76) by expansion from the bottom row.

Chapter 22

Solution 22.2: Monodromy matrix from second variations of the action. *If we take two points in the configuration space q and q' connected with a trajectory with energy E and vary them in such a way that the variation of their initial and final points are transverse to the velocity of the orbit in that point, we can write the variations of the initial and final momenta as*

$$\delta p_{\perp i} = \frac{\partial^2 S(q,q',E)}{\partial q_{\perp i}\partial q_{\perp k}}\delta q_{\perp k} + \frac{\partial^2 S(q,q',E)}{\partial q_{\perp i}\partial q'_{\perp k}}\delta q'_{\perp k} \tag{K.26}$$

and

$$\delta p'_{\perp i} = -\frac{\partial^2 S(q,q',E)}{\partial q'_{\perp i}\partial q_{\perp k}}\delta q_{\perp k} - \frac{\partial^2 S(q,q',E)}{\partial q'_{\perp i}\partial q'_{\perp k}}\delta q'_{\perp k}\,. \tag{K.27}$$

Next we express the variations of the final momenta and coordinates in terms of the initial ones. In the obvious shorthand we can write (K.27) as

$$\delta q_\perp = -S^{-1}_{q'q}S_{q'q'}\delta q'_\perp - S^{-1}_{q'q}\delta p'_\perp,$$

From (K.26) it then follows that

$$\delta p_\perp = (S_{qq'} - S_{qq}S^{-1}_{q'q}S_{q'q'})\delta q'_\perp - S_{qq}S^{-1}_{q'q}\delta p'_\perp. \tag{K.28}$$

These relations remain valid in the $q' \to q$ limit, with q on the periodic orbit, and can also be expressed in terms of the monodromy matrix of the periodic orbit. The monodromy matrix for a surface of section transverse to the orbit within the constant energy $E = H(q,p)$ shell is

$$\begin{aligned} \delta q_\perp &= \mathbf{J}_{qq}\delta q'_\perp + \mathbf{J}_{qp}\delta p'_\perp, \\ \delta p_\perp &= \mathbf{J}_{pq}\delta q'_\perp + \mathbf{J}_{pp}\delta p'_\perp. \end{aligned} \tag{K.29}$$

In terms of the second derivatives of the action the monodromy matrix is

$$\begin{aligned} \mathbf{J}_{qq} &= -S^{-1}_{q'q}S_{q'q'}\,, & \mathbf{J}_{qp} &= -S^{-1}_{q'q}\,, \\ \mathbf{J}_{pq} &= (S_{qq'} - S_{qq}S^{-1}_{q'q}S_{q'q'})\,, & \mathbf{J}_{pp} &= -S_{qq}S^{-1}_{q'q}\,, \end{aligned}$$

and vice versa

$$\begin{aligned} S_{qq} &= \mathbf{J}_{pp}\mathbf{J}^{-1}_{qp}\,, & S_{qq'} &= \mathbf{J}_{pq} - \mathbf{J}_{pp}\mathbf{J}^{-1}_{qp}\mathbf{J}_{qq}, \\ S_{q'q} &= -\mathbf{J}^{-1}_{qp}\,, & S_{q'q'} &= -\mathbf{J}^{-1}_{qp}\mathbf{J}_{qq}. \end{aligned}$$

Now do exercise 22.3.

Solution 22.3: Jacobi gymnastics. *We express the Jacobi matrix elements in* $\det(\mathbf{1}-\mathbf{J})$ *with the derivative matrices of* S

$$\det(1-J) = \det\begin{pmatrix} I + S^{-1}_{q'q}S_{q'q'} & S^{-1}_{q'q} \\ -S_{qq'} + S_{qq}S^{-1}_{q'q}S_{q'q'} & I + S_{qq}S^{-1}_{q'q} \end{pmatrix}.$$

We can multiply the second column with $S_{q'q'}$ *from the and substract from the first column, leaving the determinant unchanged*

$$\det(1-J) = \det\begin{pmatrix} I & S^{-1}_{q'q} \\ -S_{qq'} - S_{q'q'} & I + S_{qq}S^{-1}_{q'q} \end{pmatrix}.$$

Then, we multiply the second column with $S_{q'q}$ *from the right and compensate this by dividing the determinant with* $\det S_{q'q}$

$$\det(1-J) = \det\begin{pmatrix} I & I \\ -S_{qq'} - S_{q'q'} & S_{q'q} + S_{qq} \end{pmatrix} / \det S_{q'q}.$$

Finally we subtract the first column from the second one

$$\det(1-J_j)) = \det\begin{pmatrix} I & 0 \\ S_{qq'} + S_{q'q'} & S_{qq'} + S_{q'q'} + S_{q'q} + S_{qq} \end{pmatrix} / \det S_{q'q}.$$

The last determinant can now be evaluated and yields the desired result (22.15)

$$\det(1-J_j) = \det(S_{qq'} + S_{q'q'} + S_{q'q} + S_{qq}) / \det S_{q'q}.$$

Chapter ??

Solution ??: The one-disk scattering wave function.

$$\psi(\vec{r}) = \frac{1}{2}\sum_{m=-\infty}^{\infty}\left(H^{(2)}_m(kr) - \frac{H^{(2)}_m(ka)}{H^{(1)}_m(ka)}H^{(1)}_m(kr)\right)e^{im(\Phi_r-\Phi_k)}\ . \tag{K.30}$$

(For $r<a$, $\psi(\vec{r})=0$ *of course.)*

(Andreas Wirzba)

Solution ??: Ghosts do not exist. *In ref. [1] the ghost cancellation rule (??) is proved for the convolution of two* $\mathbf{A}$*-matrices and generalized to hold also inside an arbitrary (periodic) itinerary (with and without creeping sections).*

Consider the itinerary $(1,\underline{2},3,4,\underline{5},6)$ *with ghost sections at disk* 2 *and* 5 *resulting from the sixth order trace. Its geometrical contribution cancels in the trace-log expansion against the geometrical reduction of the itineraries* $(1,\underline{2},3,4,6)$, $(1,3,4,\underline{5},6)$ *from the 5th-order trace with ghost sections at disk 2 or 5, respectively, and against the geometrical reduction of the itinerary* $(1,3,4,6)$ *of the 4th-order trace with no ghost contribution:*

$$\begin{aligned}
&-\frac{1}{6}\left(6\,\underline{\mathbf{A}}^{1,\underline{2}}\underline{\mathbf{A}}^{\underline{2},3}\mathbf{A}^{3,4}\underline{\mathbf{A}}^{4,\underline{5}}\underline{\mathbf{A}}^{\underline{5},6}\mathbf{A}^{6,1}\right)\\
&\qquad-\frac{1}{5}\left(5\,\underline{\mathbf{A}}^{1,\underline{2}}\underline{\mathbf{A}}^{\underline{2},3}\mathbf{A}^{3,4}\mathbf{A}^{4,6}\mathbf{A}^{6,1}+5\,\mathbf{A}^{1,3}\mathbf{A}^{3,4}\underline{\mathbf{A}}^{4,\underline{5}}\underline{\mathbf{A}}^{\underline{5},6}\mathbf{A}^{6,1}\right)\\
&\qquad-\frac{1}{4}\left(4\,\mathbf{A}^{1,3}\mathbf{A}^{3,4}\mathbf{A}^{4,6}\mathbf{A}^{6,1}\right)\\
&=\quad(-1+2-1)\,\mathbf{A}^{1,3}\mathbf{A}^{3,4}\mathbf{A}^{4,6}\mathbf{A}^{6,1}=0\,.
\end{aligned}$$

The prefactors $-1/4$, $-1/5$, $-1/6$ *result from the trace-log expansion, the factors 4, 5, 6 inside the brackets are due to the cyclic permutations, and the rule (??) was used. If there are two or more ghost segments adjacent to each other, the ghost rule (??) has to be generalized to*

$$\begin{aligned}
&\cdots\underline{\mathbf{A}}^{i,\underline{i+1}}\underline{\mathbf{A}}^{\underline{i+1},\underline{i+2}}\cdots\underline{\mathbf{A}}^{\underline{i+k},\underline{i+k+1}}\cdots\underline{\mathbf{A}}^{\underline{i+n-1},i+n}\cdots\\
&=\quad\cdots\left(-\underline{\mathbf{A}}^{i,\underline{i+2}}\right)\cdots\underline{\mathbf{A}}^{\underline{i+k},\underline{i+k+1}}\cdots\underline{\mathbf{A}}^{\underline{i+n-1},i+n}\cdots\\
&=\quad\cdots\underline{\mathbf{A}}^{i,\underline{i+3}}\cdots\underline{\mathbf{A}}^{\underline{i+k},\underline{i+k+1}}\cdots\underline{\mathbf{A}}^{\underline{i+n-1},i+n}\cdots\\
&=\quad\cdots(-1)^{n-1}\mathbf{A}^{i,i+n}\cdots\,.
\end{aligned}\tag{K.31}$$

Finally, let us discuss one case with a repeat, e.g. the itinerary $(1,\underline{2},3,4,1,\underline{2},3,4)$ *with repeated ghost sections at disk 2 in the semiclassical limit. The cancellations proceed in the trace-log expansion as follows:*

$$\begin{aligned}
&-\frac{1}{8}\left(4\,\underline{\mathbf{A}}^{1,\underline{2}}\underline{\mathbf{A}}^{\underline{2},3}\mathbf{A}^{3,4}\mathbf{A}^{4,1}\underline{\mathbf{A}}^{1,\underline{2}}\underline{\mathbf{A}}^{\underline{2},3}\mathbf{A}^{3,4}\mathbf{A}^{4,1}\right)\\
&\qquad-\frac{1}{7}\left(7\,\underline{\mathbf{A}}^{1,\underline{2}}\underline{\mathbf{A}}^{\underline{2},3}\mathbf{A}^{3,4}\mathbf{A}^{4,1}\mathbf{A}^{1,3}\mathbf{A}^{3,4}\mathbf{A}^{4,1}\right)\\
&\qquad-\frac{1}{6}\left(3\,\mathbf{A}^{1,3}\mathbf{A}^{3,4}\mathbf{A}^{4,1}\mathbf{A}^{1,3}\mathbf{A}^{3,4}\mathbf{A}^{4,1}\right)\\
&=\quad\left(-\frac{1}{2}+1-\frac{1}{2}\right)\left[\mathbf{A}^{1,3}\mathbf{A}^{3,4}\mathbf{A}^{4,1}\right]^2=0
\end{aligned}$$

Note that the cyclic permutation factors of the 8th and 6th order trace are halved because of the repeat. The occurrence of the ghost segment in the second part of the 7th order itinerary is taken care of by the weight factor 7.

(Andreas Wirzba)

Chapter G

Solution 14.3:

(d) In the $A = 9/2$ case all cycles up to length 9 yield $\lambda = 1.08569\ldots$. (Vadim Moroz)

Solution G.1: *Using the multiplicative property of the Jacobi matrix we can write*

$$\Lambda^{t'+t}(x_0, \mathbf{u}_0) = ||\mathbf{J}^{t'+t}(x_0)\mathbf{u}_0|| = ||\mathbf{J}^{t'}(x(t))\mathbf{J}^{t}(x_0)\mathbf{u}_0||.$$

We can introduce the time evolved unit vector

$$\mathbf{u}(t) = \mathbf{J}^{t}(x_0)\mathbf{u}_0/||\mathbf{J}^{t}(x_0)\mathbf{u}_0||.$$

Then

$$||\mathbf{J}^{t'}(x(t))\mathbf{J}^{t}(x_0)\mathbf{u}_0|| = ||\mathbf{J}^{t'}(x(t))\mathbf{u}(t)||||\mathbf{J}^{t}(x_0)\mathbf{u}_0||,$$

which is the desired result.

We have to adjoin the tangent space, since the stretching factor depends on $\mathbf{u}$ *and not just on* x*. The stretching factor is multiplicative along the entire trajectory* $(x(t), \mathbf{u}(t))$*. However, it is not multiplicative along the phase space trajectory* $x(t)$ *with a fixed* $\mathbf{u}$*.*

Solution G.2: *If* $b = a^2$ *and* $T_b = 2T_a$ *we can introduce the variable* $y = e^{sT_a}$*. The dynamo rate equation then reads*

$$0 = 1 - x + x^2.$$

The solutions of this are $x_\pm = (1 \pm i\sqrt{3})/2$*. The dynamo rate is then a complex cojugate pair* $\nu = \log x_\pm / T_a$*.*

The escape rate equation is

$$0 = 1 - x/a - x^2/a^2.$$

The solutions are $x_\pm = a(-1 \pm \sqrt{5})/2$*. The escape rate is* $\gamma = \log(x_+)/T_a$*.*

In the reverse case the escape rate remains unchanged, while the dynamo rate becomes $\nu = \log((\sqrt{5}+1)/2)/T_a$*. In this case the advected field grows with an exponential rate. In the previous case it shows oscillations in addition to the exponential growth due to the imaginary part of the rate.*

Appendix L

Projects

You are urged to try 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

L.1 Deterministic diffusion, zig-zag map

To illustrate the main idea of chapter 18, 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 (18.10) 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

$$\hat{f}(x) = \begin{cases} \Lambda x & x \in [0, 1/4 + 1/4\Lambda] \\ -\Lambda x + (\Lambda+1)/2 & x \in [1/4 + 1/4\Lambda, 3/4 - 1/4\Lambda] \\ \Lambda x + (1-\Lambda) & x \in [3/4 - 1/4\Lambda, 1] \end{cases} . \tag{L.1}$$

The corresponding circle map $f(x)$ is obtained by modulo the integer part. The elementary cell map $f(x)$ is sketched in fig. L.1. The map has the symmetry property

$$\hat{f}(\hat{x}) \;=\; -\hat{f}(-\hat{x})\,, \tag{L.2}$$

so that the dynamics has no drift, and all odd derivatives of the generating function (18.3) with respect to β evaluated at $\beta = 0$ vanish.

The cycle weights are given by

$$t_p = z^{n_p} \frac{e^{\beta \hat{n}_p}}{|\Lambda_p|}\,. \tag{L.3}$$

The diffusion constant formula for 1-d maps is

$$D = \frac{1}{2} \frac{\langle \hat{n}^2 \rangle_\zeta}{\langle n \rangle_\zeta} \tag{L.4}$$

where the "mean cycle time" is given by

$$\langle n \rangle_\zeta = z \frac{\partial}{\partial z} \frac{1}{\zeta(0,z)} \bigg|_{z=1} = -{\sum}' (-1)^k \frac{n_{p_1} + \cdots + n_{p_k}}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}\,, \tag{L.5}$$

the mean cycle displacement squared by

$$\langle \hat{n}^2 \rangle_\zeta = \frac{\partial^2}{\partial \beta^2} \frac{1}{\zeta(\beta,1)} \bigg|_{\beta=0} = -{\sum}' (-1)^k \frac{(\hat{n}_{p_1} + \cdots + \hat{n}_{p_k})^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}\,, \tag{L.6}$$

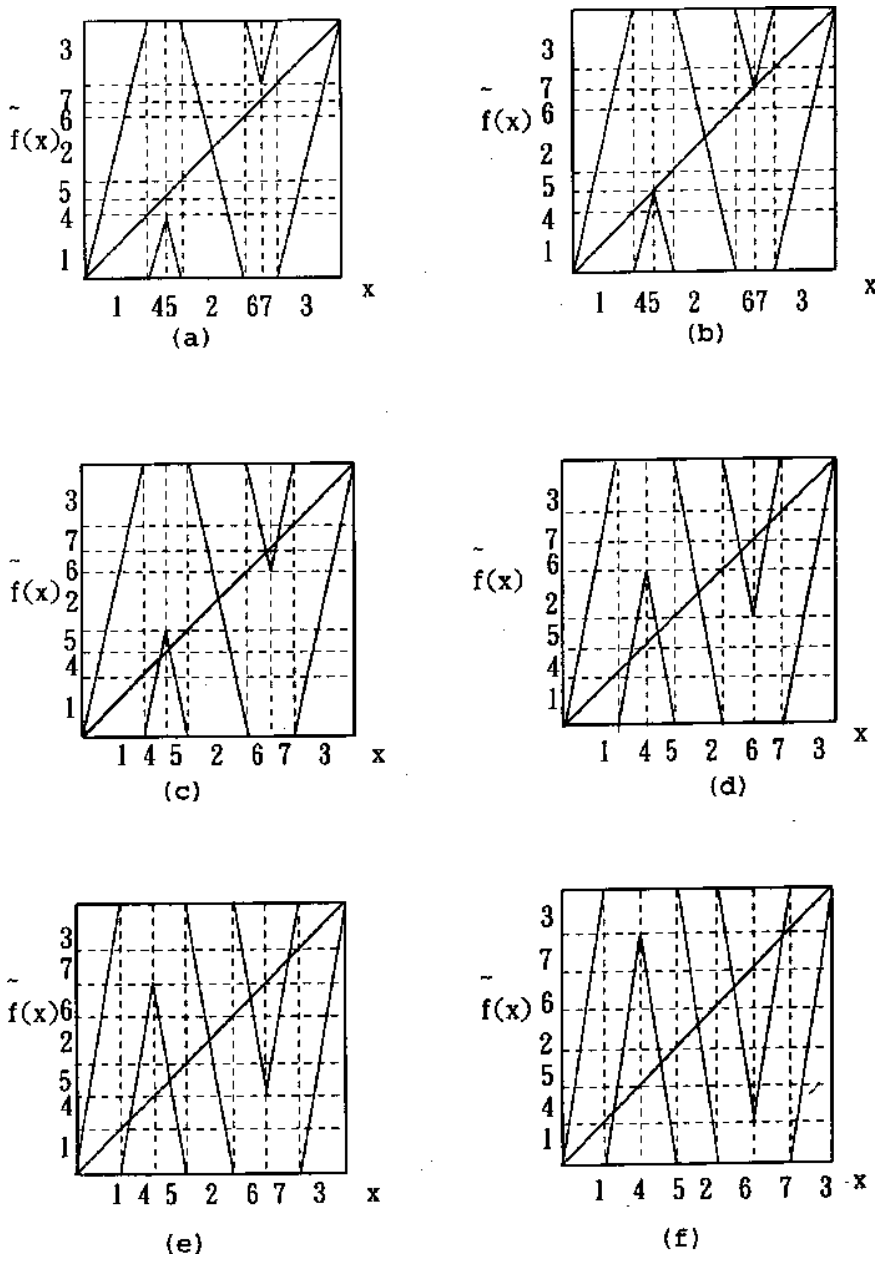


Figure L.1: (a)-(f) The sawtooth map (L.1) for the 6 values of parameter a for which the folding point of the map aligns with the endpoint of one of the 7 intervals and yields a finite Markov partition (from ref. [1]). The corresponding Markov graphs are given in fig. L.2.

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 (18.28) is convenient, but not essential for this project, so if you find sect. 17.1.2 too long a read, skip the symmetrization.

L.1.1 The full shift

Take the map (L.1) and extend it to the real line. As in example of fig. 18.3, denote by a the critical value of the map (the maximum height in the unit cell)

$$a = \hat{f}(\frac{1}{4} + \frac{1}{4\Lambda}) = \frac{\Lambda+1}{4}\,. \tag{L.7}$$

Describe the symbolic dynamics that you obtain when a is an integer, and derive the formula for the diffusion constant:

$$D \;=\; \frac{(\Lambda^2-1)(\Lambda-3)}{96\Lambda} \qquad \text{for } \Lambda = 4a-1,\ a\in\mathbb{Z}\,. \tag{L.8}$$

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 (18.20). 18.1 on p. 424

L.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 fig. L.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

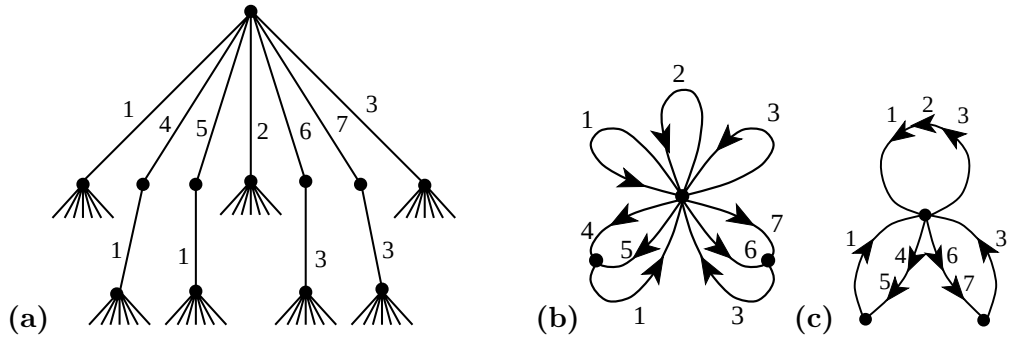


Figure L.2: (a) The sawtooth map (L.1) partition tree for fig. L.1(a); while intervals $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ map onto the whole unit interval, $f(\mathcal{M}_1) = f(\mathcal{M}_2) = f(\mathcal{M}_3) = \mathcal{M}$, intervals $\mathcal{M}_4, \mathcal{M}_5$ map onto $\mathcal{M}_1$ 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 fig. 10.14), yielding (b) the Markov graph for the Markov partition of fig. L.1(a). (c) the Markov graph in the compact notation of (18.25).

(10.2) is given by

$$\phi' = T\phi \;=\; \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_4 \\ \phi_5 \\ \phi_2 \\ \phi_6 \\ \phi_7 \\ \phi_3 \end{pmatrix} \tag{L.9}$$

and the dynamics is unrestricted in the alphabet

$$\{1,\, \underline{41},\, \underline{51},\, 2,\, \underline{63},\, \underline{73},\, 3,\,\}\,.$$

One could diagonalize (L.9) on the computer, but, as we saw in sect. 10.8, the Markov graph fig. L.2(b) corresponding to fig. L.1(a) offers more insight into the dynamics. The dynamical zeta function

$$\begin{aligned} 1/\zeta &= 1 - (t_1 + t_2 + t_3) - 2(t_{14} + t_{37}) \\ 1/\zeta &= 1 - 3\frac{z}{\Lambda} - 4\cosh\beta\frac{z^2}{\Lambda^2}\,. \end{aligned} \tag{L.10}$$

follows from the loop expansion (11.12) of sect. 11.3.

The material flow conservation sect. 14.2 and the symmetry factorization (18.28) 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 Λ. Conversely, we can use the desired Markov partition topology to write down the corresponding dynamical zeta function, and use the $1/\zeta(0,1) = 0$ condition to fix Λ. For more complicated transition matrices the factorization (18.28) is very helpful in reducing the order of the polynomial condition that fixes Λ.

The diffusion constant follows from (18.29) and (L.4)

$$\langle n\rangle_\zeta = -\left(1+\frac{1}{\Lambda}\right)\left(-\frac{4}{\Lambda}\right), \quad \langle \hat{n}^2\rangle_\zeta = \frac{4}{\Lambda^2}$$

$$D = \frac{1}{2}\,\frac{1}{\Lambda+1} = \frac{1}{10}$$

Think up other non-integer values of the parameter for which the symbolic dynamics is given in terms of Markov partitions: in particular consider the cases illustrated in fig. L.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 L.1. It is not clear why the final answers tend to be so simple. Numerically, the case of fig. L.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 L.1, fig. L.1 and (L.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 point is preperiodic to a unimodal map 3-cycle, in which case the grammar is of golden mean type, with _00_ substring prohibited (see fig. 10.14). 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.

fig. L.1	Λ	D
	3	0
(a)	4	$\frac{1}{10}$
(b)	$\sqrt{5}+2$	$\frac{1}{2\sqrt{5}}$
(c)	$\frac{1}{2}(\sqrt{17}+5)$	$\frac{2}{\sqrt{17}}$
(c')	5	$\frac{2}{5}$
(d)	$\frac{1}{2}(\sqrt{33}+5)$	$\frac{1}{8}+\frac{5}{88}\sqrt{33}$
(e)	$2\sqrt{2}+3$	$\frac{1}{2\sqrt{2}}$
(f)	$\frac{1}{2}(\sqrt{33}+7)$	$\frac{1}{4}+\frac{1}{4\sqrt{33}}$
	7	$\frac{2}{7}$

Table L.1: The diffusion constant as function of the slope Λ for the $a = 1, 2$ values of (L.8) and the 6 Markov partitions of fig. L.1

L.1.3 Diffusion coefficient, numerically

(optional:)
Attempt a numerical evaluation of

$$D = \frac{1}{2} \lim_{n\to\infty} \frac{1}{n} \left\langle \hat{x}_n^2 \right\rangle \ . \tag{L.11}$$

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 Λ; perhaps blow up a small region to show that the dependance of D on the parameter Λ is fractal. Compare with figures in refs. [1, 2, 4, 5].

L.1.4 D is a nonuniform function of the parameters

(optional:)
The dependence of D on the map parameter Λ is rather unexpected - even though for larger Λ more points are mapped outside the unit cell in one iteration, the diffusion constant does not necessarily grow. An interpretation of this lack of monotonicity would be interesting.

You can also try applying periodic orbit theory to the sawtooth map (L.1) for a random "generic" value of the parameter Λ, for example $\Lambda = 6$. The idea is to bracket this value of Λ by the nearby ones, for which higher and higher iterates of the critical value $a = (\Lambda+1)/4$ fall onto the partition boundaries, compute the exact diffusion constant for each such approximate Markov partition, and study

their convergence toward the value of D for $\Lambda = 6$. Judging how difficult such problem is already for a tent map (see sect. 11.6 and appendix E.1), this is too ambitious for a week-long exam.

References

[L.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).

[L.2] C.-C. Chen, "Diffusion Coefficient of Piecewise Linear Maps", *Phys. Rev.* **E51**, 2815 (1995).

[L.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).

[L.4] R. Klages and J.R. Dorfman, "Simple Maps with Fractal Diffusion Coefficients", *Phys. Rev. Lett.* **74**, 387-390 (1995)

[L.5] R. Klages, *Deterministic diffusion in one-dimensional chaotic dynamical systems*, (W & T, Berlin, 1996)

[L.6] R. Klages and J.R. Dorfman, "Dynamical crossover in deterministic diffusion", *Phys. Rev.* **E 55**, R1247 (1997).

L.2 Deterministic diffusion, sawtooth map

To illustrate the main idea of chapter 18, 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 (18.9). The translational symmetry (18.10) 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 fig. 18.3. The map has the symmetry property

$$\hat{f}(\hat{x}) \;=\; -\hat{f}(-\hat{x})\,, \tag{L.12}$$

so that the dynamics has no drift, and all odd derivatives of the generating function (18.3) with respect to β evaluated at $\beta = 0$ vanish.

The cycle weights are given by

$$t_p = z^{n_p} \frac{e^{\beta \hat{n}_p}}{|\Lambda_p|}\,. \tag{L.13}$$

The diffusion constant formula for 1-d maps is

$$D = \frac{1}{2} \frac{\langle \hat{n}^2 \rangle_\zeta}{\langle n \rangle_\zeta} \tag{L.14}$$

where the "mean cycle time" is given by

$$\langle n \rangle_\zeta = z \frac{\partial}{\partial z} \frac{1}{\zeta(0,z)}\bigg|_{z=1} = -{\sum}' (-1)^k \frac{n_{p_1} + \cdots + n_{p_k}}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}\,, \tag{L.15}$$

the mean cycle displacement squared by

$$\langle \hat{n}^2 \rangle_\zeta = \frac{\partial^2}{\partial \beta^2} \frac{1}{\zeta(\beta,1)}\bigg|_{\beta=0} = -{\sum}' (-1)^k \frac{(\hat{n}_{p_1} + \cdots + \hat{n}_{p_k})^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}\,, \tag{L.16}$$

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.

fig. 18.4	Λ	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 L.2: The diffusion constant as function of the slope Λ for the $\Lambda = 4, 6$ values of (18.19) and the 5 Markov partitions like the one indicated in fig. 18.4.

L.2.1 The full shift

Reproduce the formulas of sect. 18.2.1 for the diffusion constant D for Λ both even and odd integer.

L.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 (18.27) of sect. 18.2.3 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 Λ for which the symbolic dynamics is given in terms of Markov partitions: in particular consider the remaing four cases for which the critical point is mapped onto a border of a partion 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 L.2. It is not clear why the final answers tend to be so simple. What value of Λ appears to yield the maximal diffusion constant?

The 7 cases considered here (see table L.2 and fig. 18.4) 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. Some details of how this is accomplished are given in appendix **??** for a related problem, the pruned Bernulli shift.

L.2.3 Diffusion coefficient, numerically

(optional:)
Attempt a numerical evaluation of

$$D = \frac{1}{2} \lim_{n\to\infty} \frac{1}{n} \left\langle \hat{x}_n^2 \right\rangle \ . \tag{L.17}$$

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 Λ; perhaps blow up a small region to show that the dependance of D on the parameter Λ is fractal. Compare with figures in refs. [1, 2].

L.2.4 D is a nonuniform function of the parameters

(optional:)
The dependence of D on the map parameter Λ is rather unexpected - even though for larger Λ more points are mapped outside the unit cell in one iteration, the diffusion constant does not necessarily grow. Fig. 18.5 taken from ref. [1] 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 (18.9) for a random "generic" value of the parameter Λ, for example $\Lambda = 4.5$. The idea is to bracket this value of Λ by the nearby ones, for which higher and higher iterates of the critical value $a = \Lambda/2$ 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. 11.6 and appendix E.1), this is too ambitious for a week-long exam.

References

[L.1] R. Klages and J.R. Dorfman, "Simple Maps with Fractal Diffusion Coefficients", *Phys. Rev. Lett.* **74**, 387-390 (1995)

[L.2] R. Klages, *Deterministic diffusion in one-dimensional chaotic dynamical systems*, (W & T, Berlin, 1996)

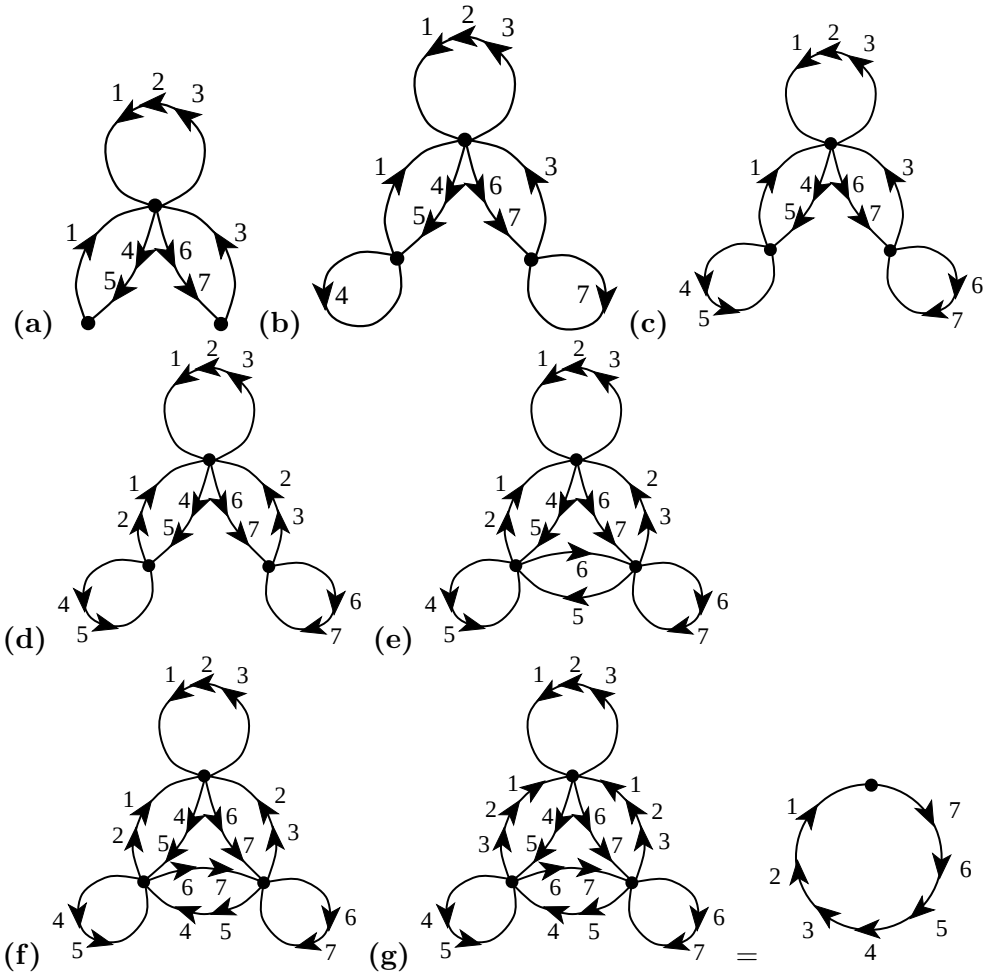


Figure L.3: (a)-(g) Markov graphs corresponding to the cases of fig. L.1 for which the sawtooth map (L.1) folding point aligns with an endpoint of one of the 7 intervals and yields a finite Markov partition. In the case (g) the large Markov graph is equivalent to the simple unrestricted shift in 7 symbols. Graphs (f), (e), ..., are obtained by pruning links from the unrestricted expanded graph (g).

Chapter L

Solution, project sect. L.1: Markov partitions for the zig-zag map

Fig. L.1(b) $\Lambda = \sqrt{5}+2$ with the corresponding Markov graph given in fig. L.3(b). As the 4 and 5 links (and other sets of links) are parallel, we have simplified the graph using the compact notation of (18.25). The loop expansion (11.12) yields

$$\begin{aligned} 1/\zeta &= 1-(t_1+t_2+t_3)-(t_4+t_7) \\ &\quad -2(t_{14}+t_{37})+(t_1+t_2+t_3)(t_4+t_7)+t_4t_7 \end{aligned}$$

$$
\begin{aligned}
& +2(t_{14}t_7 + t_4t_{37}) - (t_1+t_2+t_3)t_4t_7 \\
= \;& 1 - 3\frac{z}{\Lambda} - 2\cosh\beta\frac{z}{\Lambda} + \frac{z^2}{\Lambda^2} + 2\cosh\beta\frac{z^2}{\Lambda^2} + \frac{z^3}{\Lambda^3}\,.
\end{aligned} \tag{L.18}
$$

The Markov graphs are not unique, and extra terms cancel in the evaluation. Flow conservation (14.11) and symmetry factorization (**??**) yield

$$
0 = 1/\zeta(0,1) = \left(1 - \frac{1}{\Lambda}\right)\left(1 - \frac{4}{\Lambda} - \frac{1}{\Lambda^2}\right)
$$

The second factor indeed vanishes for the given value of Λ. The diffusion constant follows from

$$
\langle n\rangle_\zeta = -\left(1-\frac{1}{\Lambda}\right)\left(-\frac{4}{\Lambda} - \frac{2}{\Lambda^2}\right), \quad \langle \hat{n}^2\rangle_\zeta = \frac{2}{\Lambda} - \frac{4}{\Lambda^2} = \frac{2}{\Lambda}\left(1-\frac{2}{\Lambda}\right)
$$

$$
D = \frac{1}{2\sqrt{5}}
$$

Fig. L.1(c) $\Lambda = (\sqrt{17}+5)/2$ with the corresponding Markov graph given in fig. L.3(c).

$$
\begin{aligned}
1/\zeta &= 1 - (t_1+t_2+t_3) - 2(t_4+t_7) \\
&\quad -2(t_{14}+t_{37}) + 2(t_1+t_2+t_3)(t_4+t_7) + 4t_4t_7 \\
&\quad +4(t_{14}t_7 + t_4t_{37}) - 4(t_1+t_2+t_3)t_4t_7 \\
&= 1 - 3\frac{z}{\Lambda} - 4\cosh\beta\frac{z}{\Lambda} + 4\frac{z^2}{\Lambda^2} + 8\cosh\beta\frac{z^2}{\Lambda^2} - 4\frac{z^3}{\Lambda^3}\,,
\end{aligned} \tag{L.19}
$$

Flow conservation (14.11) and symmetry factorization (**??**) yield

$$
0 = 1/\zeta(0,1) = \left(1 - \frac{2}{\Lambda}\right)\left(1 - \frac{5}{\Lambda} + \frac{2}{\Lambda^2}\right)
$$

The second factor indeed vanishes for the given value of Λ. The diffusion constant follows from

$$
\langle n\rangle_\zeta = -\left(1-\frac{2}{\Lambda}\right)\left(-\frac{5}{\Lambda} - \frac{4}{\Lambda^2}\right), \quad \langle \hat{n}^2\rangle_\zeta = \frac{4}{\Lambda} - \frac{8}{\Lambda^2} = \frac{2}{\Lambda}\left(1-\frac{2}{\Lambda}\right)
$$

$$
D = \frac{2}{\sqrt{17}}
$$

It is not clear why the final answers tend to be so simple. Numerically this is the maximal diffusion constant. Is it? Why?

Fig. L.1(d) $\Lambda = (\sqrt{33}+5)/2$ with the corresponding Markov graph given in fig. L.3(d).

$$\begin{aligned} 1/\zeta &= 1-(t_1+t_2+t_3)-2(t_4+t_7) \\ &\quad -4(t_{14}+t_{37})+2(t_1+t_2+t_3)(t_4+t_7)+4t_4t_7 \\ &\quad +8(t_{14}t_7+t_4t_{37})-4(t_1+t_2+t_3)t_4t_7 \\ &= 1-3\frac{z}{\Lambda}-4\cosh\beta\frac{z}{\Lambda}+4\frac{z^2}{\Lambda^2}+4\cosh\beta\frac{z^2}{\Lambda^2}+4\frac{z^3}{\Lambda^3}\,. \end{aligned} \tag{L.20}$$

Flow conservation (14.11) and symmetry factorization (**??**) yield

$$0 = 1/\zeta(0,1) = \left(1-\frac{2}{\Lambda}\right)\left(1-\frac{5}{\Lambda}-\frac{2}{\Lambda^2}\right)$$

The second factor indeed vanishes for the given value of Λ. The diffusion constant follows from

$$\langle n\rangle_\zeta = -\left(1-\frac{2}{\Lambda}\right)\left(-\frac{5}{\Lambda}-\frac{4}{\Lambda^2}\right)\,, \quad \langle \hat{n}^2\rangle_\zeta = \frac{4}{\Lambda}-\frac{4}{\Lambda^2} = \frac{4}{\Lambda}\left(1-\frac{1}{\Lambda}\right)$$

$$D = \frac{1}{8}+\frac{5}{88}\sqrt{33}$$

Fig. L.1(e) $\Lambda = 2\sqrt{2}+3$ with the corresponding Markov graph given in fig. L.3(e).

$$\begin{aligned} 1/\zeta &= 1-3t_1-2(t_4+t_7) \\ &\quad -4(t_{14}+t_{37})-t_{47}+6t_1(t_4+t_7)+4t_4t_7 \\ &\quad -4(t_{147}+t_{347})+8(t_{14}t_7+t_4t_{37})+3t_1t_{47}-12t_1t_4t_7 \\ &= 1-3\frac{z}{\Lambda}-4\cosh\beta\frac{z}{\Lambda}+3\frac{z^2}{\Lambda^2}+4\cosh\beta\frac{z^2}{\Lambda^2}-\frac{z^3}{\Lambda^3}\,. \end{aligned} \tag{L.21}$$

Markov graphs are not unique, and the extra terms cancel in the evaluation. Flow conservation (14.11) and symmetry factorization (**??**) yield

$$0 = 1/\zeta(0,1) = \left(1-\frac{1}{\Lambda}\right)\left(1-\frac{6}{\Lambda}+\frac{1}{\Lambda^2}\right)$$

The second factor indeed vanishes for the given value of Λ. The diffusion constant follows from

$$\langle n\rangle_\zeta = -\left(1-\frac{1}{\Lambda}\right)\left(-\frac{6}{\Lambda}+\frac{2}{\Lambda^2}\right), \quad \langle \hat{n}^2\rangle_\zeta = \frac{4}{\Lambda}-\frac{4}{\Lambda^2} = \frac{4}{\Lambda}\left(1-\frac{1}{\Lambda}\right)$$

$$D = \frac{1}{2\sqrt{2}}$$

Fig. L.1(f) $\Lambda = (\sqrt{33}+7)/2$ with the corresponding Markov graph given in fig. L.3(f). Go through the whole construction, check that:

$$\begin{aligned} 1/\zeta &= 1-3t_1-2(t_4+t_7) \\ &\quad -4(t_{14}+t_{37}+t_{47})+6t_1(t_4+t_7)+4t_4t_7 \\ &\quad -8(t_{417}+t_{437})+8(t_{14}t_7+t_4t_{37})+12t_1t_{47}-12t_1t_4t_7 \\ &= 1-3\frac{z}{\Lambda}-4\cosh\beta\frac{z}{\Lambda}+4\cosh\beta\frac{z^2}{\Lambda^2}\,. \end{aligned} \tag{L.22}$$

Flow conservation (14.11) and symmetry factorization (**??**) yield

$$0 = 1/\zeta(0,1) = \left(1-\frac{7}{\Lambda}+\frac{4}{\Lambda^2}\right)$$

The diffusion constant follows from

$$\langle n\rangle_\zeta = \left(-\frac{7}{\Lambda}+\frac{8}{\Lambda^2}\right) = \left(\frac{7}{\Lambda}-2\right), \quad \langle \hat{n}^2\rangle_\zeta = \frac{4}{\Lambda}-\frac{4}{\Lambda^2} = -\frac{3}{\Lambda}+1$$

$$D = \frac{1}{4}+\frac{1}{4\sqrt{33}}$$

Fig. L.1(g) $\Lambda = 7$. The Markov graph corresponding to fig. L.1(g) is fig. L.3(g).

$$\begin{aligned} 1/\zeta &= 1-3t_1-2(t_4+t_7) \\ &= 1-3\frac{z}{\Lambda}-4\cosh\beta\frac{z}{\Lambda} \end{aligned} \tag{L.23}$$

The diffusion constant follows from (L.8)

$$\langle n\rangle_\zeta = \frac{7}{\Lambda}, \quad \langle \hat{n}^2\rangle_\zeta = \frac{4}{\Lambda}$$

$$D = \frac{2}{7}$$